Beyond the threshold: theoretical and empirical nonlinear time-series econometrics of foreign exchange markets

Beyond the Threshold:
Theoretical and Empirical Nonlinear Time-Series Econometrics of Foreign Exchange Markets

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Universiteit Maastricht
op gezag van Rector Magnificus, Prof. mr. G.P.M.F Mols,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen
op woensdag 29 juni 2005 om 16:00 uur

door

Michel René van Tol

Promotor:
Prof. dr. C.C.P. Wolff

Beoordelingscommissie:
Prof. dr. P.C. Schotman (voorzitter)
Prof. dr. P.H.B.F. Franses (Erasmus Universiteit Rotterdam)
Dr. J.R.Y.J. Urbain

ISBN 90 9019534 3

Preface

99% percent of Finance people give the rest a bad name.

Part of that 1% are, undoubtedly, my colleagues and friends at the University of Maastricht. This dissertation would never have come to be without the concerted help of many others. Be it not on technical matters it would be through social connectedness. For this and the many enjoyable outings that we made with the department I am very grateful.

I would like to thank all of my colleagues at the Finance Department at the University of Maastricht but especially to Christian, Stefan, Stefanie, Thorsten, Anja, Rachel, Rob and Boris. Boris, I'll miss your jokes without a punch line and the many funny stories. I would also like to thank the ladies of the secretariat without you I would have been even more lost than usual. I would like to include a special thanks to Margy: I will certainly miss our deeply personal discussions and you as a confident in difficult times.

I would really like to express my gratitude to Christian: I could not have hoped for a better promoter. The freedoms you afforded me have surely shaped the way that I now define myself.

To all my friends: Mus, Ivo, Sara, Hakan, Till, Pieter, Ralf, Corine and Sydney. Thanks for the totally amazing times we had together. Tinna thank you so much for your patience and love.

Oma, Papa, Mama, Beverly, Eduard, Tandi, Kyle and Damian: Thanks for everything.

Michel R. van Tol
Maastricht, May 2005

Contents

List of Tables

List of Figures

1

Introduction

1.1 Introduction

The seminal work of Meese and Rogoff (1983), poignantly, exemplified the incapacity of monetary models of exchange rate determination, developed and utilized during the 1970's, in adequately modeling the mechanisms that underlie the generation of exchange rates. As a consequence of their findings two broad developments, in the exchange rate forecasting literature, may be discerned. The first delineates an area of research that attempts to expound on extant monetary models, based, broadly speaking, on macroeconomic lines of argumentation, while the second details the advancement of econometric time-series based exchange rate models. This dissertation is an attempt at contributing to the growing body of research which has resulted through the unstinted efforts of many researchers at reconciling these two developments.

In this dissertation we develop new econometric theory, designed to further advance the reconciliation alluded to previously, with the intention of ultimately applying it to the task of

forecasting exchange rates. As such, we detail a progression from a critical note on the application of linear cointegration techniques to the task of forecasting spot exchange rates via the information content of forward premiums; to the exposition of a new fitting approach for multiple-threshold Threshold Autoregressive, or TAR, models; which is, subsequently, utilized in the development of a nonlinear unit root test; and the presentation of a novel fitting approach in multivariate threshold cointegrated processes. The contribution of these developments is then assessed by juxtaposing an empirical study, based on linear cointegration, with an equivalent study utilizing these techniques.

1.2 Motivation

Numerous researchers, over the past two decades, have sought to study the relative performance of structural exchange rate models vis-à-vis atheoretical time series models (see for example, Meese and Rogoff (1983, 1986); Boughton (1987); Booth and Glassman (1987)). Empirical evidence is generally suggestive of the fact that structural models are only capable of marginal improvements in out-of-sample forecasting for monthly and quarterly exchange rates relative to a random walk model. The usefulness of structural models is frequently constrained by the sampling frequency of many of the input variables, such as relative prices, inflation rates, etc., which are at best measured on a monthly basis. As a result, at higher frequencies, given that structural models are ruled out, our attention will need to be directed towards the plausible alternative of time series modeling.

There is some evidence that financial market participants utilize price histories when making predictions of their future values. Taylor and Allen (1992), having surveyed major dealers in the foreign exchange market in London, found that, at short horizons of one week or less, 90% of respondents used some form of chartism, with 60% stating that they regarded the information as at least as important as economic fundamentals. These findings are in congruence with numerous other studies that find strong support for the use of technical trading rules, from the viewpoint of their widespread application in financial markets (Frankel and Froot (1990)) to their capacity to produce surprisingly accurate results (Pruitt and White (1988, 1989); Brock *et al.* (1992)). Additional support for the latter result was found by Neely *et al.* (1997) who, using a genetic programming technique to find technical trading rules, found strong evidence of economically significant out-of-sample excess returns to those rules while finding no evidence that these returns were compensation for bearing systematic risk.

Clyde and Osler (1997) were able to make the important link between technical trading rules and non-linear forecasting; showing that technical analysis can generate higher profits than a random trading strategy if the true data generating process[1] is not linear[2]. Collectively, these observations form a strong motivation for the consideration of nonlinear time series models of price histories for forecasting financial asset prices or returns.

It has generally been accepted that nonlinearity is a stylized fact of financial market returns. Hinich and Patterson (1985), Scheinkman and LeBaron (1989), Hsieh (1989,1993), Mayfield and Mizrach (1992) and Pfann *et al.* (1996), for example, find strong evidence of nonlinearity in numerous asset return series. Hsieh (1989) and Brooks (1996) were able to show that significant nonlinearity remains in a series after allowing for volatility clustering effects, the feature to which most of the nonlinear behavior is typically attributed.

From a more technical perspective the basis for linear modeling approaches, such as that of Box and Jenkins, is the Wold Representation, which states that *any* covariance stationary time-series can be expressed as a moving average function of present and past innovations. While this infinite moving average can nearly always be well approximated by low order autoregressive processes Granger (1983) in his article "Forecasting White Noise" illustrated that a lack of autocorrelation in a time series does not imply a lack of predictability of the series. More strongly, Brock and Chamberlain (1984) illustrated how some perfectly predictable time-series have zero autocorrelations at all lags. Additionally, Taylor (2001) showed that a linear model gives biased results towards a random walk model if the true model is nonlinear with a substantial "band of inaction". Resultantly, given the empirically pervasive finding that the random walk model dominates alternative empirical models in terms of predictability across financial markets, nonlinear time-series modeling appears to be a logical endeavor.

Three types of nonlinear models have commonly been adopted in the literature: Markov Switching Models (see Hamilton (1989)); Threshold Autoregressive Models (see Tong (1983, 1990)); and Smooth Transition Autoregressive Models (see Chan and Tong (1986)). In this dissertation we focus on models of the Threshold Autoregressive-type or more specifically on Self-Exciting Threshold Autoregressive, or SETAR, models. For many the abrupt regime changes characteristic of these models is unrealistic while the non-standard likelihood/least

[1] The existance of a true data generating process is still being debated as is evidenced by Hansen's (2005) comment on the dialogue between Granger and Hendry, in which he identifies the assumption of a true data generating process as a conceptual error in the theory of econometric model selection.

[2] One of the most prevelant stylized facts of exchange rates is the fat-tailedness of their empirical distributions implying their generation by a process that is not linear.

squares functions are frequently viewed as a distraction. Compounding these perspectives is the view that the econometric theory underlying STAR models is more eloquent than its equivalent for SETAR models, recent theoretical advances are debasing that argument: see, amongst others, Hansen (1997) on inference in TAR models; Caner and Hansen (1997), Enders and Granger (1998), Berben and van Dijk (1999) and Kapetanios and Shin (2002, 2003a, 2003b) on unit root testing; and Medeiros *et al.* (2002) and Coakley *et al.* (2003) for estimation approaches.

A frequently cited advantage when choosing to utilize smooth-transition autoregressive, or STAR, models, which nest the SETAR model class, is the availability of a reparameterization when the value of the delay parameter is known which permits the circumvention of the so-called Davies problem. In the general case, however, to which our interest in this dissertation extends, when the value of the delay parameter needs to be estimated this problem remains.

Recent empirical evidence suggests that both real and nominal exchange rates exhibit SETAR-like behavior. Kapetanios and Shin (2003a) using a newly developed test of nonstationary long memory against nonlinear ergodic models find that the yen real exchange rates, versus a very large number of countries, points (robustly) in the direction of SETAR nonlinearity. The results when specifying a STAR model under the alternative are significantly less attractive since the test is unable to distinguish between long memory and nonlinearity for all values of the delay parameter. A possible explanation for this finding, provided by the authors, is that SETAR-type nonlinearity is more likely to result when markets are liquid than when they are not; stating that the more liquid the market the greater the possibility of sudden adjustments. Theoretically, it is apparent that models that predict smooth behavior for the real exchange rate, do not rule out the possibility of discontinuous adjustment as a limit case, hence these findings do not necessarily contradict the predictions of these models.

Evidence that similar behavior is found in nominal rates, at least in the context of a target zone, is provided by Vilasuso and Cunningham (1996) who test for nonlinearity of EMS exchange rates. Their findings suggest that these rates were linear in the period 1979 through 1987, being consistent with the realignment target zone model, whereas many currencies conformed to a nonlinear process during the succeeding period up to 1992, implying a shift towards the credible target zone model. Additional support is provided by Bessec (2003) in which she developed a target zone model where intramarginal interventions occur beyond a band of inaction, being consistent with the (Krugman (1991)) credible target zone model, and

was able to show that central bank intervention dynamics were consistent with this model. Cuaresma *et al.* (2004) report similar findings for an extended set of currencies on a more recently sampled dataset. Statistically, these findings imply SETAR-type behaviour[3] or, more specifically, the existence of Band-TAR like exchange rate dynamics.

Studies of currency crises suggest that central banks are frequently more willing to and capable of resisting pressure for currency appreciation than depreciation (see for example Flood and Marion (1998) and the references therein). This asymmetry is generally explained by the higher costs of resisting depreciation pressure than appreciation pressure, these costs are usually measured in terms of sacrifices of objectives, other than that of exchange rate stabilization, pursued by a central bank, and may concern unemployment, competitiveness, economic growth, inflation and/or the viability of financial institutions due to its role as a lender of last resort (see Obstfeld (1990)).

Cumulatively, these studies suggest that nominal and real exchange rates, measured on a weekly/monthly basis, behave nonlinearly with discontinuous adjustment and therefore provide the motivation for our interest in this particular class of models in this dissertation.

1.3 Outline

The essence of this dissertation comprises four chapters with a distinct contentual progression. To provide a context for our point of departure we adumbrate these chapters with an introductory overview of exchange rate processes and relevant econometric techniques in Chapter 2.

It is notoriously difficult to out-predict a simple random walk model in out-of-sample forecasting exercises of foreign exchange rate processes. Hence, when a study claims to outperform the random walk, at various horizons, it is unreservedly considered with some skepticism. Chapter 3 is a comment on the work of Clarida and Taylor (1997) who develop a framework for the prediction of future spot rates by extracting information from the term structure of forward exchange premiums. They do so by showing that while both the spot and forward rates are individually integrated processes an economically justifiable combination of these processes is stationary leading naturally to their disentanglement through the use of

[3] A plausible explanation may be provided by sampling frequency. One could conjecture the existence of smooth transition dynamics at high sampling frequencies, for a given process, while discontinuous adjustment dominates at lower frequencies.

linear cointegration techniques. Utilizing the same data and methodology we were unable to corroborate their findings i.e. we find significantly less statistical support for the empirical framework purported by the authors than they claim, even when taking into account potential reasons for the discrepancy. The empirical framework on which their study hinges is interesting however and leads quite naturally to the idea of including nonlinear short-run dynamics in the equilibrating process of a cointegrated system; a notion introduced into the literature by Balke and Fomby (1997).

Threshold cointegration relies on the estimation of TAR models. Consequently, Chapter 4 details a new fitting approach, based on the fitting approach of Coakley *et al.* (2003), for estimating the parameters of a multiple-threshold TAR model. The study presents a novel grid construction algorithm that takes into account typically instituted requirements, as well as, the formulation needed for the simple implementation of updating algorithms. Based on this grid we partition the continuous threshold space into a set of non-overlapping hypercubes, each of which is sequentially activated and the residual sum of squares function, which Coakley *et al.* (2003) have proven is a rational function of a specific order, is subjected to a newly developed interpolation-based optimization routine. Monte Carlo simulations indicate the relative superiority of the proposed approach in terms of estimation time. Summarily, the approach is 2, 32, 570 and 14 000 times more expedient at estimating 3-, 4-, 5- and 6- regime TAR models than the conventional procedure[4].

Chapter 5 delineates a direct testing procedure of the unit root hypothesis; specifying stationary double-threshold momentum-TAR models under the alternative. One reason for doing so relates to the findings of Cook (2003) which indicate that practitioners are more likely to detect momentum-TAR rather than TAR-type nonlinearity when modeling an unknown process. The test is comprehensive in that we allow for a lag-augmented data generating mechanism, as well as, a range of values for the threshold delay variable. Additionally, the critical values that we present allow for the presence of either/both constant and trend attractors. Power and size tests of the proposed testing procedure indicate that the test is superior to a pertinent set of alternative tests in terms of power yet is prone to over-rejection; these size distortions are not uncommon however. We apply the tests in a study of the term structure of UK Eurocurrency interest rates. There we illustrate that the interest rate differential is stationary and best described by a double-threshold Band-MTAR model. Given these findings and the presence of a single unit root in the interest rates themselves, we

[4] While we exposit the results of the three and four regime models we only generate conservative estimates of the estimation time of 5 and 6 regime models by linear extrapolation.

estimate a bivariate threshold vector error correction model, or TVECM, and provide strong evidence that if the threshold behavior of the interest rate differential is not accounted for a "linear" VECM is significantly misspecified.

Chapter 6 introduces an estimation routine for multivariate threshold cointegrated systems. Surprisingly, while recent developments in the literature have focused on bivariate threshold cointegrated systems the more general multivariate case has been left untended. We present a convenient formulation of a multivariate threshold VECM and suggest a two-step procedure, as is typical in the simpler bivariate case, by estimating (SE)TAR models of the cointegrating residual series and conditioning parameter estimates of the multivariate threshold VECM on these estimates. We apply these constructs in a reinvestigation of the empirical framework of Clarida and Taylor (1997), and the comment of the third chapter. Interestingly, we are able to show strong support for the presence of single-threshold nonlinearity in the forward premiums and, hence, the necessity of incorporating these characteristics into the cointegration framework implied by their empirical derivations. Out-of-sample forecasting exercises indicate the relative superiority of the multivariate threshold VECM approach to the standard "linear" VECM. An important finding is that despite the inclusion of the threshold behavior or asymmetry in the equilibrium reversion process we are still unable to systematically out-predict the random walk model. Finally, our results indicate that the model does very well in the short-run but deteriorates strongly at longer forecasting horizons.

Chapter 7 concludes with a brief summary of the results and their implications. Finally, we discuss a series of suggestions for future research.

2

Time Series Modeling of Foreign Exchange Processes – A Theoretical Overview

2.1 Introduction

Prior to Slutztky's (1927) observation, that stable low-order linear stochastic difference equations could generate cyclic processes, nonlinear models played an important role in modeling economic dynamics. The ensuing espousal of the "linear paradigm" in economic research was not unfounded however; the empirical success of models based on stochastic difference equations was substantial and hence justifiable. Recent studies have shown that key macroeconomic variables display asymmetric adjustment over the course of the business cycle, however, precipitating in the current interest in nonlinear models.

These developments are pervasive in economics, including foreign exchange economics, to which our interest in this dissertation extends. In this chapter we detail recent theoretical and empirical advances related to foreign exchange modeling; primarily to lend a degree of

comprehensibility to the piece by elucidating the econometric fundaments that will form our point of departure.

2.2 Foreign Exchange Forecasting

The demise of the post-war Bretton Woods system of fixed exchange rates, in 1973, led most of the large industrialized economies to float their exchange rates. Ensuing research has largely focused on the development and estimation of empirical models of floating exchange rates; primarily for purposes of providing a sound foundation for economic policy decisions. Floating exchange rates are, however, notoriously difficult to forecast at short- to medium-term horizons. Monetary models of exchange rate determination, such as Dornbusch's "overshooting" theory, do possess some explanatory power at longer horizons, while short horizon forecasts, based on standard models, utilizing observable macroeconomic fundamentals, are typically readily outperformed by a (driftless) random walk. The latter was poignantly exposited by Meese and Rogoff (1983a,b) who showed how a simple martingale process is able to outperform more complex structural models – in forecasting exchange rates up to a year in advance – despite providing the structural models with *ex post* information on future fundamentals such as money and output. Attempts at overturning their results have typically led to corroborative conclusions. A sideline result of this work is the standard adoption of the simple random walk model as a benchmark when evaluating empirical exchange rate models.

Advances in the empirical investigation of exchange rate processes have led to a series of pervasive results that are consistent with the following, non-exhaustive list, of propositions (see Mussa (1979)):

1. The log of the spot exchange rate is approximately a random walk;
2. Most changes in exchange rates are unexpected;
3. The currencies of countries with high inflation rates tend to depreciate, and at approximately the inflation differential in the long-run;
4. Both the nominal exchange rate and fundamentals, such as ratios of money to real output, appear to be propelled by a unit root, while real exchange rates are typically stationary.

5. Actual exchange rate movements appear to overshoot movements in smoothly adjusting equilibrium exchange rates[5].
6. The relationship between monetary fundamentals and the nominal exchange rate does not appear to be close in the short-run, though they may share longer-run trends.
7. The assumption of flexible prices is contradicted, by the considerable correlation between nominal and real exchange rates at high frequencies while appearing to fall at lower frequencies, indicating the possible presence of price stickiness.
8. The real exchange rate is contemporaneously related to the real interest differential[6].

The apparent failure, beyond that of theoretical coherence, of fundamentals-based structural exchange rate models in outperforming a driftless random walk led to an inspired burst of research investigating univariate characteristics of nominal exchange rates. An interesting characteristic is the stabilizing speculation that Friedman (1953) argued would dominate exchange rate dynamics under a floating regime. When expectations are such that an appreciation today – relative to some long-run equilibrium – is to induce market participants to forecast depreciation in the future they are said to be stabilizing. Frankel and Froot (1987a) found stabilizing speculation dominated the expectations of investors in the medium-term, while a plethora of research (see, amongst others, Frankel and Froot (1987b, 1990), Froot and Ito (1989) and Ito (1994)) seems to indicate that the opposite takes place at short horizons.

Compounding the innate complexity of exchange rate processes is a variety of other factors one example of which is the, statistically verifiable findings, that psychological aspects, such as barriers at round numbers (see De Grauwe and Decupere (1992)) also affect exchange rate processes. The promising field of behavioral finance, which attempts, to find structure in the often less than fully rational decisions of economic agents, by focusing on investor psychology, is still in an embryonic stage of development yet has, nevertheless, led to the general consensus that this new perspective is likely to contribute to our understanding of the movements made by asset prices, including those of exchange rates.

---

[5] The classical theoretical paper on short-run price stickiness is due to Dornbusch (1976) in which he illustrated how nominal exchange rates may "overshoot" their long-run levels.

[6] While Meese and Rogoff (1988) utilizing the Engle-Granger test for cointegration are unable to verify the existence of such a relation a large number of subsequent studies have found evidence to the contrary (see, for example, Blundell-Wignall and Browne (1991), Throop (1993) and Baxter (1994).

The rest of this chapter is devoted to the task of introducing the class of (SE)TAR models and their extension into tests of the unit root and threshold cointegration hypotheses.

2.3 Non-linear Modeling of Time Series Processes

2.3.1 TAR Models

A time series z_t is said to be a k-regime TAR process if it follows the model

$$z_t = \sum_{j=1}^{k}\left[\left(\phi_0^{(j)} + \sum_{l=1}^{p^{(j)}}\phi_l^{(j)} z_{t-l}\right) I_t\left(\theta^{(j-1)} \le v_{t-d} < \theta^{(j)}\right)\right] + \varepsilon_t \tag{2.1}$$

where $j, d, p^{(j)} \in \aleph$, $I_t(A)$ is an indicator function such that it equals unity when the event A occurs and is zero otherwise, $\phi_l^{(j)}$ denotes the autoregressive coefficient for the *lth* lag in regime j and v_{t-d} denotes the delay variable. If $v_{t-d} \equiv z_{t-d}$ then (2.1) is usually referred to as a self-exciting TAR (SETAR) model, typically referred to with relevant parameter settings i.e. SETAR$\left(p^{(1)}, p^{(2)}, \ldots, p^{(k)}; d; k\right)$. The k linear AR regimes are determined by comparing the delay variable with the $k+1$ vector of threshold values, $\boldsymbol{\theta} = \left[\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(k)}\right]$, where $\theta^{(0)} = -\infty$ and $\theta^{(k)} = \infty$. Finally, ε_t is assumed to be a martingale difference sequence with respect to the past history of z_t.

Two important variants of (2.1) are the momentum-TAR (M-TAR) and continuous TAR (CTAR) models introduced into the literature by Enders and Granger (1998) and Chan and Tsay (1998), respectively. The details of these models will be elucidated upon as they become pertinent to the discussion. Sufficient conditions to ensure ergodicity in processes represented by (SE)TAR models are presented in the next section.

2.3.2 Stationarity and Mixing Properties

Under the assumption that ε_t is i.i.d.$\left(0, \sigma_\varepsilon^2\right)$, independent of z_0, with a distribution that is absolutely continuous and positive everywhere Bec *et al.* (2004) provide a sufficient condition for ergodicity of the Markov process $X_t = \left(z_t, z_{t-1}, \ldots, z_{t-m+1}\right)$, where

$m = \max(p, d)$, in the three-regime form of (2.1)[7]. Geometric ergodicity has the following implications:

1. The existence of a unique stationary distribution of X_t;
2. X_t becomes stationary exponentially fast when initialized from an arbitrary value;
3. Once initialized from its stationary distribution X_t is stationary and $\beta-$mixing with geometric decay;
4. z_t is $\beta-$mixing with geometric decay implying that the series may be locally nonstationary, when allowing inner-regimes to be driven by a unit root/explosive process, while being globally ergodic.

If $m > p$ define the $m \times m$ matrix $\mathbf{A}_j$ as

$$\mathbf{A}_j = \begin{bmatrix} \phi_1^{(j)} & \phi_2^{(j)} & \cdots & \phi_p^{(j)} & 0 & \cdots & 0 \\ 1 & 0 & 0 & & & & \vdots \\ 0 & 1 & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & 0 & 0 \\ 0 & \cdots & \cdots & \cdots & 0 & 1 & 0 \end{bmatrix} \tag{2.2}$$

otherwise,

$$\mathbf{A}_j = \begin{bmatrix} \phi_1^{(j)} & \phi_2^{(j)} & \cdots & \cdots & \phi_p^{(j)} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & \ddots & 1 & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix} \tag{2.3}$$

and let $\|\cdot\|_{M_1}$ be a matrix norm in the sense of Horn and Johnson(1985, page 290). Then Bec *et al.* (2004) prove the following theorem:

[7] Note that the formulation of theorem 2.1, despite being specific to the three-regime case, is easily extended to the more general *k*-regime context.

Theorem 2.1 (Condition for geometric ergodicity (Theorem 1, Bec *et al.* (2004))):
Suppose that $\{z_t\}$ *satisfies (2.1) and assume that* ε_t *is* i.i.d.$(0,\sigma_\varepsilon^2)$, *independent of* z_0, *with a distribution that is absolutely continuous and positive everywhere. Assume that there exists a value of t such that*

(a) for $d \geq p$,

$$\max_{\pi_{1i},\pi_{2i}} \left\| \prod_{i=1}^{m} \mathbf{A}_1^{\pi_{1i}} \mathbf{A}_1^{\pi_{2i}} \mathbf{A}_1^{1-\pi_{1i}-\pi_{2i}} \right\|_{M_1} < 1$$

with $\sum_{i=1}^{m} \pi_{2i} < m$, *where* $\pi_{1i} = 1$ or 0, $\pi_{2i} = 1$ or 0, *and* $\pi_{1i}\pi_{2i} = 0$.

(b) for $d < p$, *besides (a) above, the following is true*

$$\left\| \mathbf{A}_2^{p} \right\|_{M_1} \leq 1.$$

If, moreover, $\{z_t\}$ *is initialized from its stationary distribution, then* $\{z_t\}$ *is stationary* $\alpha-$ *and* $\beta-$*mixing with geometric decay.*

The analysis of stationarity and ergodicity of a (SE)TAR[8] model requires that ε_t be "turned off", in (2.1), and investigating the 'skeleton' or deterministic component of the model. Chan and Tong *(1985)* illustrated that asymptotic stability is ensured when $\max_j \sum_{i=1}^{p} \left|\phi_i^{(j)}\right| < 1$, which corresponds with the requirement that the AR processes in the outer-regimes are stationary (Note that this is a much stronger than the conditions of Bec *et al.* 2004). Necessary and sufficient conditions for a TAR(p) are currently not well understood. Conditions for the asymptotic stability of the more restrictive case of a first-order model, with $p^{(j)} = 1$, for all $j = 1,2,\ldots,k$ and $d = 1$, have been investigated by Chan *et al.* (1985) and Kapetanios and Shin (2002). Stationarity, in the aforementioned case, is guaranteed if and only if one of the following conditions hold:

[8] Specific conditions for the stationarity of M-TAR processes have, as yet, received an inadequate amount of academic attention to warrant exposition at this point.

$$
\begin{array}{lll}
(i) & \phi_1^{(1)} < 1, \phi_1^{(k)} < 1 & \text{and} \quad \phi_1^{(1)}\phi_1^{(k)} < 1; \\
(ii) & \phi_1^{(1)} < 1, \phi_1^{(k)} = 1 & \text{and} \quad \phi_0^{(k)} > 0; \\
(iii) & \phi_1^{(1)} = 1, \phi_1^{(k)} < 1 & \text{and} \quad \phi_0^{(1)} < 0; \\
(iv) & \phi_1^{(1)} = \phi_1^{(k)} = 1 & \text{and} \quad \phi_0^{(1)} < 0 < \phi_0^{(k)}; \\
(v) & \phi_1^{(1)}\phi_1^{(k)} = 1, \phi_1^{(k)} < 1 & \text{and} \quad \phi_0^{(1)} + \phi_1^{(1)}\phi_0^{(k)} > 0.
\end{array}
\tag{2.4}
$$

The presence of (SE)TAR-type process dynamics has implications for conventional tests of the unit root and cointegration hypotheses, which we discuss in succeeding sections.

2.4 Unit Root Testing in Time Series Analysis

2.4.1 Introduction

Many observed time-series processes exhibit random wandering behavior; such as is typical of interest rates and asset prices. Before 1970, modeling nonstationarity entailed capturing secular movements in a series by using deterministic trending functions, such as time polynomials. More complex trend functions such as time polynomials with sinusoidal factors and piecewise time polynomials were also used. The latter corresponding with the general class of models, to which our interest in this work extends, since they allow for the presence of regime-specific behavior. The nonstochastic nature of these trending functions forms a major limitation when investigating economic processes. Models with a stochastic trend have time dependent variances that go to infinity through time; have shocks that are persistent in the sense that they have a permanent effect on the values generated by the process; and have an infinite spectrum at the origin.

Typically, testing for the presence of stochastic trend is done by utilizing

$$y_t = h_t + y_t^s \tag{2.5}$$

where $h_t = \gamma' x_t$ and $y_t^s = \alpha y_{t-1}^s + u_t$. y_t^s is a stationary time-series and x_t is a k-vector of deterministic trends, entails testing the null hypothesis that the autoregressive parameter $\alpha = 1$. A multitude of such tests exist based on parametric, semi-parametric and non-parametric methods employing both classical and Bayesian principles of statistical testing. As

presaged our concern, in this work, relates to further developing the non-linear unit root tests purported by Enders and Granger (1998), as well as, Berben and van Dijk (1999).

Conventional unit root tests are misspecified when the stochastic process underlying a particular variable of interest exhibits asymmetric reversion to a long-run equilibrium; the conjoint conclusions of the studies undertaken by Pippenger and Goering (1993), Enders and Granger (1998) and Berben and van Dijk (1999). Tests specifying a null hypothesis of a unit root against the alternative of stationarity with threshold adjustment have consequently received decided attention. The supremum-Wald unit root testing methodology when a symmetric 3-regime SETAR model is specified under the alternative hypothesis, detailed in articles by Enders and Granger (1998) and Berben and van Dijk (1999), is the point from which we depart in chapter 5[9].

2.4.2 The Sup-Wald Unit Root Testing Methodology

Suppose that model (2.1) has been compactly reformulated as

$$\Delta z = \rho_1 (z_{t-1} - \theta) I(z_{t-1} \leq \theta) + \rho_3 (z_{t-1} - \theta) I(z_{t-1} > \theta) + \varepsilon_t \tag{2.6}$$

where z_t is assumed to be a "demeaned" symmetric 3-regime SETAR (1;1;3) process, with $p = p^{(1)} = p^{(2)} = p^{(3)}$; specifies a driftless random walk as the AR process in the inner-regime; and $\rho_i = \left(\phi_1^{(i)} - 1\right)$ for $i = 1,3$. Then, when (2.6) is globally stationary, it may be specified under the alternative hypothesis in a test of the unit root hypothesis by utilizing a conventional Wald-test[10].

If the locations of the thresholds are known *a priori* then the F-statistic may be computed as

[9] Andrews and Ploberger (1994) also detail the use of the average and the exponential average of the Wald statistic. Kapetanios and Shin (2002) find the exponential average of the Wald statistic to have the most power of the three when a three-regime SETAR model is specified under the alternative. We abstain from using these statistics choosing rather to focus on the conventionally adopted supremum-Wald statistic.

[10] Unit root tests based on the stochastic multiple structural-break model developed by Bai and Perron (1998a, 1998b) appear to resemble tests based on SETAR models. Two obvious differences in the testing methodologies are that, firstly, while the Bai and Perron model entails searching for the presence of a set of structural breaks in the (discrete) time dimension SETAR-based tests, particularly those to which our interest in this dissertation extends i.e. Band-SETAR models, require a search over a (continuous) space defined in the levels dimension of the process under consideration. Secondly, SETAR models are a special case of the Bai and Perron models whereby a set of structural breaks correspond with a particular threshold level; in that sense the Bai and Perron models nest the SETAR class. In Chapter 5 we detail the connection between the two model classes more elaborately.

$$F_n(\theta) = \frac{\sum_{t=1}^{n} \Delta z_t z_{t-1}(\theta)' \left[\sum_{t=1}^{n} z_{t-1}(\theta) z_{t-1}(\theta)' \right]^{-1} \sum_{t=1}^{n} z_{t-1}(\theta) \Delta z_t}{2\hat{\sigma}_n^2(\theta)} \tag{2.7}$$

where $z_{t-1}(\theta) = \left((z_{t-1} - \theta) I(z_{t-1} \le \theta), (z_{t-1} - \theta) I(z_{t-1} > \theta) \right)'$ and $\hat{\sigma}_n^2(\theta) = \frac{1}{n} \sum_{t=1}^{n} \hat{\varepsilon}_t(\theta)^2$ where the residuals are defined as $\hat{\varepsilon}_t(\theta) = \Delta z_t - z_{t-1}(\theta)' \hat{\rho}(\theta)$. Using a drifting threshold construct Berben and van Dijk (1999) obtain the following reformulation of (2.7)

$$F_n(\theta(\tau)) = \frac{n}{2} \left(\frac{\tilde{\sigma}_n^2 - \hat{\sigma}_n^2(\theta(\tau))}{\hat{\sigma}_n^2(\theta(\tau))} \right) \tag{2.8}$$

where $\tilde{\sigma}_n^2 = \sum_{t=1}^{n} \Delta z_t$ is an estimate of the residual variance under the null hypothesis; $\theta(\tau) = (1-\tau) \min_{1 \le t \le n}(z_{t-1}) + \tau \max_{1 \le t \le n}(z_{t-1})$ where $\tau \in (0,1)$ identifies the location of the threshold as a linear combination of the minimum and maximum value of the delay variable z_{t-1}.

The drifting threshold construct is simply a one-to-one mapping from the unit interval to the continuous threshold space, thus an estimate $\hat{\tau}$ of τ corresponds with the threshold estimate $\hat{\theta}(\hat{\tau})$ of the population threshold location, the supremum-Wald statistic, when the location of the threshold is unknown, is then defined as

$$F_n \equiv F_n\left(\hat{\theta}(\hat{\tau})\right) = \sup_{\tau \in \mathrm{T}} F_n(\theta(\tau)) \tag{2.9}$$

where $\mathrm{T} = \{\theta \mid \min(z_{t-1}) \le \theta \le \max(z_{t-1})\}$. Andrews (1993) shows the importance of restricting the set of viable threshold candidates away from the maximum and minimum value of the delay variable, concluding with the recommendation that the first and last 15% of the order statistics of the delay variable be excluded. It is clear that (2.9) corresponds with the interpretation of evaluating the null hypothesis given the most likely location of the threshold. Given the central role played by unit root tests in cointegration analysis it is clear that if a process is discernibly non-linear, in a SETAR sense, linear cointegration tests are misspecified. In the sections that follow we begin by discussing linear cointegration and then introduce Balke and Fomby's (1997) concept of threshold cointegration, which is more appropriate when reversion in the disequilibrium process is decidedly discontinuous.

2.5 Cointegration Analysis

2.5.1 Linear Cointegration

If two variables x_t and y_t are $I(1)$ and there exists a linear combination of these variables that is stationary, then x_t and y_t are said to be cointegrated, a concept introduced into the literature by Granger (1981). Generally, if both x_t and y_t are $I(d)$ then they are cointegrated if, for some β, $y_t - \beta x_t \sim I(d-b)$ with $b > 0$. Hence, if two variables are cointegrated a long-run equilibrium relationship will exist between them, conversely, if they are not cointegrated regressing x_t and y_t will lead to a spurious regression, as addressed by Yule (1926). Consider the following system of equations

$$\begin{aligned} x_t &= u_t, & u_t &= u_{t-1} + \varepsilon_{1t} \\ y_t + \alpha x_t &= v_t, & v_t &= \rho v_{t-1} + \varepsilon_{2t} \end{aligned} \tag{2.10}$$

from which it is clear that x_t and y_t are $I(1)$ regardless of the value of ρ. If $|\rho| < 1$, however, then $y_t + \alpha x_t \sim I(0)$ implying that x_t and y_t are cointegrated. Suppose the existence of k variables $y_t \sim I(1)$ and $\beta' y_t = u_t \sim I(0)$, then β is termed a cointegrating vector, while, the equation $\beta' y_t = u_t$ is called the cointegrating regression[11]. Assume that two cointegrating vectors β_1 and β_2 exist so that both $\beta_1' y_t = u_{1t}$ and $\beta_2' y_t = u_{2t}$ are stationary, then, since linear combinations of $I(0)$ variables are also $I(0)$, any linear combination of the vectors is also a cointegrating vector. This result implies the existence of an identification problem.

The presaged discussion assumes, implicitly, that adjustment to the long-run equilibrium, given the stationarity of the error term, is continuous. Economic arguments are plentiful as to why this need not necessarily be the case. Transaction costs, institutional constraints and arbitrage, amongst others, often imply the existence of a threshold in the disequilibrium process, such that only once the disequilibrium exceeds a particular threshold level will reversion to the long-run equilibrium of the process occur. Understanding the need for this non-linear extension of conventional (linear) cointegration techniques Balke and Fomby (1997) developed the concept of threshold cointegration.

[11] Note that the cointegrating vector may not be a k-vector of zeros.

2.5.2 Threshold Cointegration

The amalgamation of nonlinear time-series and cointegration analysis has lead to a plenitude of models which are distinctly adept at capturing the complex dynamics characteristic of many economic phenomena, including those of exchange rates. An eloquent concept introduced into the literature by Balke and Fomby (1997) is that of threshold cointegration; whereby the disequilibrium process is well modeled by a model of the SETAR class. Threshold cointegrated processes are well represented by a class of models that allow the short-run components of variables to have a flexible non-linear dynamic specification while the long-run components obey equilibrium constraints. Consider the following simple bivariate system

$$y_t + \beta x_t = B_t, \quad B_t = B_{t-1} + \eta_t \tag{2.11}$$

$$y_t + \alpha x_t = z_t, \quad z_t = \phi_1^{(j)} z_{t-1} + \varepsilon_t, \tag{2.12}$$

with

$$\phi_1^{(j)} = \begin{cases} 1 & if \quad |z_{t-1}| \le \theta \\ \phi, \text{ with } |\phi| < 1 & if \quad |z_{t-1}| > \theta \end{cases} \tag{2.13}$$

where ε_t and η_t are assumed to be i.i.d. white noise processes and, notationally, $\phi_1^{(j)}$ is consistent with (2.1). Utilizing the Engle-Granger two-step single equation method will yield a superconsistent OLS estimate of the cointegration vector, in the first step, that is conditioned upon when estimating the parameters of the ECM, in the second step. In equation (2.11) B_t represents the common stochastic trend ensuring that x_t and y_t are $I(1)$; equation (2.12) represents the equilibrium relationship with a cointegrating vector given by $(1, \alpha)$; $|\phi| < 1$ ensures that the equilibrium error is stationary. The corresponding error correction representation may subsequently be written as

$$\Delta y_t = \gamma_1^{(j)} z_{t-1} + \upsilon_{1t} \tag{2.14}$$

$$\Delta x_t = \gamma_2^{(j)} z_{t-1} + \upsilon_{2t} \tag{2.15}$$

where

$$\gamma_1^{(J)} = -(1-\phi_1^{(J)})\beta/(\beta-\alpha);\ \gamma_2^{(J)} = (1-\phi_1^{(J)})/(\beta-\alpha);$$
$$\upsilon_{1t} = [\beta/(\beta-\alpha)]\varepsilon_t - [\alpha/(\beta-\alpha)]\eta_t;\ \upsilon_{2t} = [1/(\beta-\alpha)](\eta_t - \varepsilon_t);$$
$$z_{t-1} = y_{t-1} + \alpha x_{t-1}.$$

The authors proceed by stipulating a number of interesting SETAR models, restricted variants of model (2.1), that govern the dynamics of the equilibrium error process z_t. Two of these models, are the Equilibrium-TAR and Band-TAR models, reformulated, here, in first-difference form as

Equilibrium-TAR: $$\Delta z_t = ((\phi-1)z_{t-1})I_t(|z_{t-1}| > \theta) + \varepsilon_t \tag{2.16}$$

Band-TAR: $$\Delta z_{t-1} = ((\phi-1)(z_{t-1}-\theta))I_t(|z_{t-1} > \theta|) + \varepsilon_t \tag{2.17}$$

where $\varepsilon_t \sim IIN(0,\sigma_\varepsilon^2)$ and Δ denotes the first-difference operator. Notationally, the remainder may be inferred from (2.1). These models are particularly interesting as they have an inner-regime that is driven by a random walk; implying that stationarity is purely determined by the outer-regime processes. Importantly, allowing the disequilibrium process to be propelled by a unit root during certain sub-periods indicates a lack of cointegration during those periods and the presence thereof when the discrepancies exceed a threshold value. The benefits of being able to model behavior of this type in empirical economics are both abundant and self-evident; one clear application would be investigating the behavior of arbitrage processes in the presence of transaction costs. The advantages of threshold cointegration are counterbalanced, however, by the complexity of having to estimate a cointegrating vector in a nonstandard manner; attributable to the presence of a threshold parameter of unknown value.

2.6 Conclusion

By juxtaposing the adumbration of §2.2 with the econometric constructs of the remainder of this chapter it is evident that a significant proportion of the stylized facts of foreign exchange rates may be explained by reverting back to a "nonlinear paradigm". Engaging in nonlinear

modeling does come at the cost of computational complexity while the benefits are still to be born out by intensive empirical testing in the future.

3

Comment on "*The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors*"

3.1 Introduction

Since Meese and Rogoff's (1983) illustration that existent structural exchange rate models are unable to out-predict the naïve alternative of a driftless random walk, researchers have been applying a plethora of new econometric techniques in an attempt at uncovering empirical regularities that have thus far eluded them. Despite the rapid evolution of these techniques research subsequent to the Meese and Rogoff papers has failed to convincingly overturn their findings. On occasion statistical models are able to capture the dynamics of exchange rate movements succinctly but are usually later shown to be too sample specific and therefore poor predictors out-of-sample. Thus, when an apparently theoretically sound model produces out-of-sample forecasts that significantly out-predict the random walk, at various horizons, it

piques the interest of any who have an interest in the field. Just such results arose in an interesting investigation of the forecastability of spot exchange rates based on the information content of the term structure of forward premia by Clarida and Taylor (1997) (henceforth CT). This comment entails the empirical re-implementation of the framework set forth in their article. We show that the framework is unable to produce out-of-sample forecasts that are consistently superior to those of a naïve random walk and point to a number of possible causes for the deviations observed between results. For the sake of brevity we will not provide a detailed discussion of the framework but will state the essential details of the model under consideration and then directly address the issue of the empirical results.

3.2 Empirical Preliminaries

We obtained the Harris Bank database, maintained by Richard Levich[12], which allows us to investigate discretely sampled, weekly data on spot and 4-, 13-, 26-, and 52-week forward dollar exchange rates for Germany, Japan, and the United Kingdom. Our data span the period 1977:1 through 1993:52 and are identical to those used in CT and thus should not in principle be a source of any statistical discrepancy. All estimations are carried out using data for the period 1977:1 through 1990:26 and we reserve all complimentary data for the purpose of evaluating forecasts. After converting all data to logarithmic form we performed a number of preliminary Augmented Dickey-Fuller tests to test for unit roots in the stochastic processes underlying the sampled realizations of the spot and forward rates. In accordance with the conclusions drawn by the authors we are unable to reject the hypothesis of a unit root for the spot and forward rates, while we were able to do so for the forward premia.

3.3 The Vector Error Correction model

Under the assumption that deviations from the risk-neutral efficient markets hypothesis (RNEMH), at all horizons, are stationary stochastic processes the authors posit the existence of a cointegrating relationship between spot- and forward exchange rates. According to the

[12] We thank Richard Levich for kindly making the data available.

Granger Representation Theorem (1987) the system should therefore be well represented by a VECM[13] and can subsequently be written as

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma} \Delta \mathbf{y}_{t-1} + \boldsymbol{\Pi} \mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{3.1}$$

where $\mathbf{y}_t = \left[s_t, f_{4,t}, f_{13,t}, f_{26,t}, f_{52,t} \right]'$ is a vector of spot- and forward rates at time t. Further, if cointegration between the variables in the system exists, $\boldsymbol{\Pi}$ will be a matrix of reduced rank and may be decomposed into two matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ such that $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$.

3.4 Tests of Cointegration Rank

The authors postulate the existence of a single common trend propelling the dynamics of the five-variable system. Stock and Watson's (1988) result, that a test for cointegration rank (r) in an n-variable system is equivalent to a test for $n-r$ common trends, implies that we need to show that the system has a cointegration rank of 4. Table I presents the estimates of the λ-max and Trace statistics used to determine cointegration rank. The results indicate that in accordance with the findings of the article the dollar-sterling and dollar-yen systems are cointegrated with rank 4 at the 5% critical level. The first significant departure between results occurs in the dollar-Mark system where we find that we can only be 90% certain that it is cointegrated with rank 4. The difference in results may be attributable to data manipulation be it through the correction of typographical errors, the removal of extreme changes in the data or incorrectly inputting the data into the statistical algorithm[14,15]. Here we assume that the relatively simple statistical algorithm was implemented correctly. These deviations will have an effect on the FIML estimates of the VECM and subsequently we expect that the forecasts generated by our model are different from that of CT.

[13]Due to instability of the full information maximum-likelihood algorithm the authors restricted the dimensionality of the lag space of the VECM to one. Since we are attempting to replicate the results conditional on the decisions made by the authors we implemented an equivalent restriction.

[14] We opted to leave the data in its original form after consulting Richard Levich on the presence of known errors in the data set. Despite this we did correct one clear typographical error in the dollar-sterling set.

[15] We performed the same test using other statistical packages and came to the same conclusion i.e. given the data the statistics that we calculate must be correct.

Table I

Tests of Cointegrating Rank of $\mathbf{y}_t = [s_t, f_{4,t}, f_{13,t}, f_{26,t}, f_{52,t}]$

Sample period is 1977:1 to 1990:26. Critical values are from Osterwald – Lenum (1992, Table 1)

	λ-max Statistic	10% Critical Level	Trace Statistic	10% Critical Level
Dollar-sterling				
H_0: $r \leq 4$	4.773	2.69	4.773	2.69
H_0: $r \leq 3$	16.664	12.07	21.438	13.33
Dollar-mark				
H_0: $r \leq 4$	0.824	2.69	0.824	2.69
H_0: $r \leq 3$	12.255	12.07	13.079	13.33
Dollar-yen				
H_0: $r \leq 4$	1.387	2.69	1.387	2.69
H_0: $r \leq 3$	20.523	12.07	21.910	13.33

3.5 The Economic Hypothesis

The uninterpretable cointegration space is replaced, through a set of parametric restrictions on the cointegrating vectors, by the economic hypothesis of interest. A likelihood ratio test is then performed to test whether the parametric restrictions significantly alter the maximized likelihood. Assuming the existence of four cointegrating vectors the likelihood ratio statistic is distributed as $\chi^2(4)$ under the null. The results are reported in Table II.

Table II

Tests of the Null Hypothesis that Four Linearly Independent Forward Premiums Comprise a Basis for the Cointegration Space.

Sample period is 1977:1 to 1990:26. The test is conditional on there being four linearly independent cointegrating vectors.

	$\chi^2(4)$	Marginal Significance Level
Dollar-mark	6.92	14.00%
Dollar-yen	13.65	0.85%
Dollar-sterling	22.78	0.01%

We find that we are unable to reject the hypothesis that the forward premiums define a basis for the cointegration space for the dollar-mark. We do however reject the hypothesis for the dollar-sterling and the dollar-yen exchange rates. These conclusions differ from those reported in the article and subsequently we find significantly less support for the economic hypothesis implicit in the empirical framework. At this point we come to a juncture where we either reject the model outright on the basis of a lack empirical support or we continue the analysis because the authors did so and produced the results that sparked our interest in the first place. Thus, despite the lack of support we continue with the analysis under the assumption that the four linearly independent forward premiums comprise a basis for the cointegration space.

3.6. Results

3.6.1 FIML Error Correction Model Results

Conditional on a Gaussian error distribution for the VECM, we use full information maximum likelihood to derive parameter estimates as is described in Johansen (1995). The results are reported in Tables III, IV, and V[16]. We decided to report all estimated parameters and not exclude those variables that had less significant coefficients; a result of being unable to locate a significance level that would result in an estimated model with the variables reported in CT.

We find that, other than for the dollar-mark system, the coefficients of the estimated models have no clearly discernable pattern, though given the relatively high associated standard errors we are unable to draw a concrete conclusion in this regard. Multicollinearity between independent variables is likely to be the reason for the high uncertainty of the coefficient estimates. The Box-Ljung and Hosking's portmanteau statistics indicate a lack of inter-temporal correlation in the residual series in both a univariate and multivariate sense. In fact, these measures of model adequacy deviate significantly in magnitude to those reported by CT; this may however be a result of the exclusion restrictions that the authors implemented. The differences in the estimated models are as yet indiscernible hence by subjecting the models to an out-of-sample forecasting exercise we hope to bring statistically important differences to the foreground.

[16] CT do not indicate which data series they use for the spot rate. We decided to use the average of the high and low spot rates during a week as our variable s_t of interest. We also performed the analysis on the closing bid price, the lowest spot rate during the week and the highest spot rate during the week. The models estimated do vary but the out-of-sample performance is relatively consistent in that there is no model that out-predicts the random walk at any horizon.

Table III
FIML Error Correction Model for the Five-Variable System: Dollar-Sterling

Sample period is 1977:1 to 1990:26. The Q-statistics are Box-Ljung statistics computed at 13 autocorrelations of the residual series; H is Hosking's multivariate portmanteau statistic computed at 13 autocorrelations; All distributions are distributed as central χ^2 under the null hypothesis, with the degrees of freedom indicated. Figures in parentheses are marginal significance levels.

	Model for Δs_t		Model for Δs_t		Model for Δs_t		Model for Δs_t		Model for Δs_t	
Regressor	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Δs_{t-i}	-0.071	0.936	-0.087	0.935	-0.336	0.939	-0.390	0.957	-0.678	0.994
$\Delta f_{4,t-i}$	0.323	1.216	0.262	1.215	0.498	1.221	0.390	1.244	0.464	1.292
$\Delta f_{13,t-i}$	0.363	0.757	0.440	0.756	0.452	0.760	0.633	0.774	0.668	0.804
$\Delta f_{26,t-i}$	-0.819	0.650	-0.823	0.650	-0.830	0.653	-0.864	0.665	-0.368	0.691
$\Delta f_{52,t-i}$	0.375	0.278	0.376	0.278	0.387	0.279	0.400	0.284	0.088	0.295
$(s-f_4)_{t-i}$	-0.168	0.983	-0.387	0.982	-0.095	0.986	0.263	1.005	-0.343	1.044
$(s-f_{13})_{t-i}$	0.149	0.141	-0.381	0.140	-0.403	0.141	-0.414	0.144	-0.376	0.149
$(s-f_{26})_{t-i}$	0.262	0.087	0.263	0.086	0.265	0.087	0.291	0.088	0.329	0.092
$(s-f_{52})_{t-i}$	-0.031	0.210	-0.027	0.2101	0.026	0.211	0.022	0.215	0.080	0.223
Constant	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001	-0.001	0.001
	Q(13) = 12.70 (0.47)		Q(13) = 12.71 (0.47)		Q(13) = 12.66 (0.47)		Q(13) = 11.84 (0.54)		Q(13) = 12.03 (0.52)	
					H(325) = 388.33 (0.01)					

Table IV
FIML Error Correction Model for the Five-Variable System: Dollar-Mark

Sample period is 1977:1 to 1990:26. The Q-statistics are Box-Ljung statistics computed at 13 autocorrelations of the residual series; H is Hosking's multivariate portmanteau statistic computed at 13 autocorrelations; All distributions are distributed as central χ^2 under the null hypothesis, with the degrees of freedom indicated. Figures in parentheses are marginal significance levels.

	Model for Δs_t		Model for Δs_t		Model for Δs_t		Model for Δs_t		Model for Δs_t	
Regressor	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Δs_{t-i}	-0.348	1.299	-0.382	1.295	-0.432	1.292	-0.653	1.294	-0.908	1.307
$\Delta f_{4,t-i}$	1.250	1.733	1.312	1.729	1.409	1.724	1.633	1.727	1.789	1.745
$\Delta f_{13,t-i}$	0.097	1.044	0.047	1.041	-0.060	1.039	0.043	1.040	0.071	1.051
$\Delta f_{26,t-i}$	-1.207	0.634	-1.216	0.632	-1.263	0.631	-1.489	0.632	1.241	0.638
$\Delta f_{52,t-i}$	0.397	0.333	0.429	0.332	0.529	0.332	0.646	0.332	0.475	0.335
$(s-f_4)_{t-i}$	-2.028	1.420	-2.168	1.416	-1.400	1.412	-2.062	1.415	-1.567	1.429
$(s-f_{13})_{t-i}$	-1.227	0.783	-1.307	0.781	-1.155	0.779	-0.791	0.780	-1.265	0.788
$(s-f_{26})_{t-i}$	-0.441	0.223	-0.476	0.222	-0.523	0.222	-0.562	0.222	-0.562	0.224
$(s-f_{52})_{t-i}$	0.133	0.079	0.132	0.078	0.135	0.078	0.143	0.078	0.162	0.079
Constant	0.004	0.001	0.004	0.001	0.004	0.001	0.004	0.001	0.004	0.001
	Q(13) = 19.33 (0.11)		Q(13) = 19.76 (0.10)		Q(13) = 19.85 (0.10)		Q(13) = 18.97 (0.10)		Q(13) = 17.89 (0.16)	
					H(325) = 464.54 (0.00)					

Table V
FIML Error Correction Model for the Five-Variable System: Dollar-Yen

Sample period is 1977:1 to 1990:26. The Q-statistics are Box-Ljung statistics computed at 13 autocorrelations of the residual series; H is Hosking's multivariate portmanteau statistic computed at 13 autocorrelations; All distributions are distributed as central χ^2 under the null hypothesis, with the degrees of freedom indicated. Figures in parentheses are marginal significance levels.

Regressor	Model for Δs_t		Model for Δs_t		Model for Δs_t		Model for Δs_t		Model for Δs_t	
	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE	Coeff.	SE
Δs_{t-1}	-0.989	0.817	-0.903	0.819	-0.909	0.820	-0.960	0.823	-0.881	0.835
$\Delta f_{4,t-1}$	2.587	1.086	2.537	1.088	2.552	1.089	2.486	1.094	2.318	1.110
$\Delta f_{13,t-1}$	-0.495	0.872	-0.602	0.874	-0.700	0.874	-0.558	0.878	-0.798	0.891
$\Delta f_{26,t-1}$	-1.211	0.636	-1.151	0.637	-1.100	0.638	-1.129	0.641	-0.609	0.650
$\Delta f_{52,t-1}$	0.311	0.274	0.322	0.275	0.353	0.275	0.351	0.276	0.157	0.280
$(s-f_4)_{t-1}$	0.316	1.078	-0.054	1.081	0.655	1.082	0.255	1.086	0.335	1.102
$(s-f_{13})_{t-1}$	-0.416	0.427	-0.525	0.428	-0.529	0.428	-0.351	0.430	-0.675	0.436
$(s-f_{26})_{t-1}$	0.141	0.083	0.139	0.083	0.144	0.083	0.164	0.084	0.209	0.085
$(s-f_{52})_{t-1}$	-0.166	0.057	-0.173	0.057	-0.183	0.057	-0.189	0.058	-0.185	0.058
Constant	-0.004	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001	-0.004	0.001
	Q(13) = 15.31 (0.29)		Q(13) = 14.77 (0.32)		Q(13) = 14.42 (0.34)		Q(13) = 14.33 (0.35)		Q(13) = 14.35 (0.35)	
					H(325) = 487.45 (0.00)					

3.6.2 Out-of-Sample Forecast Results

We generate a series of spot rate forecasts dynamically up to 52 weeks ahead over the period 1990:27 to 1993:52. The methodology employed is identical to that of the original article. We report the RMSE and the MAE in levels for the estimated VECMs and then calculate a ratio statistic for four alternative models including a driftless random walk[17]. A ratio of less than unity indicates the relative superiority of the VECM over an alternative model. The results are reported in Tables VI, VII, and VIII.

[17] The alternative models include an unrestricted forth order VAR in the five series, a forward premium regression, and a standard forward rate model.

Table VI
Results of the Forecasting Exercises: Dollar-Sterling

Notes: Forecast period is 1990:27 to 1993:52. For the VECM the RMSE and the MAE is expressed in levels. For the alternative forecasts, the RMSE or the MAE is expressed as the inverse of its ratio to the corresponding figure for the VECM. Thus a figure less than 1 indicates the relative superior performance by the VECM.

	VECM (level)	VAR (ratio)	Random Walk (ratio)	Forward Premium Regression (ratio)	Forward Rate (ratio)
Root-mean-square error (RMSE)					
4-week horizon	0.0396	1.007	1.022	0.982	1.025
13-week horizon	0.0883	0.664	1.163	0.998	1.191
26-week horizon	0.1393	0.115	1.346	1.006	1.431
52-week horizon	0.2060	0.001	1.763	1.074	2.067
Mean absolute error (MAE)					
4-week horizon	0.0291	0.987	1.000	0.985	1.002
13-week horizon	0.0633	0.630	1.143	0.988	1.163
26-week horizon	0.0984	0.113	1.396	1.007	1.476
52-week horizon	0.1782	0.002	2.074	1.042	2.404

The VECM produces satisfactory forecasts as is evidenced by relatively small forecast errors but is nevertheless unable to significantly out-predict the naïve random walk at any horizon. The obvious difference in results lies in the development of the forecasting performance as the forecasting horizon increases in length. As opposed to the findings of the authors we find that the VECM performs progressively worse versus the random walk the farther into the future we look. A finding that is consistent with our expectations and is congruent with the patterns found in complementary research in the field of exchange rate prediction (see Wolff (1987); Wolff (1988); and Meese and Rogoff (1988)). We are well aware that the estimated models are designed to take a long-term equilibrium into account and thus should perform relatively well for predictions at longer horizons. As expected this pattern is present when comparing the VECM to the unrestricted VAR but not when compared to any of the other alternative models.

Table VII
Results of the Forecasting Exercises: Dollar-Mark

Notes: Forecast period is 1990:27 to 1993:52. For the VECM the RMSE and the MAE is expressed in levels. For the alternative forecasts, the RMSE or the MAE is expressed as the inverse of its ratio to the corresponding figure for the VECM. Thus a figure less than 1 indicates the relative superior performance by the VECM.

	VECM (level)	VAR (ratio)	Random Walk (ratio)	Forward Premium Regression (ratio)	Forward Rate (ratio)
Root-mean-square error (RMSE)					
4-week horizon	0.0351	1.032	1.046	0.979	1.047
13-week horizon	0.0826	0.803	1.254	1.036	1.266
26-week horizon	0.1354	0.195	1.614	1.051	1.644
52-week horizon	0.2091	0.002	2.550	1.151	2.676
Mean absolute error (MAE)					
4-week horizon	0.0271	1.010	1.016	0.962	1.017
13-week horizon	0.0638	0.788	1.198	1.014	1.211
26-week horizon	0.1047	0.195	1.682	1.032	1.749
52-week horizon	0.1831	0.003	3.177	1.122	3.252

Table VIII
Results of the Forecasting Exercises: Dollar-Yen

Notes: Forecast period is 1990:27 to 1993:52. For the VECM the RMSE and the MAE is expressed in levels. For the alternative forecasts, the RMSE or the MAE is expressed as the inverse of its ratio to the corresponding figure for the VECM. Thus a figure less than 1 indicates the relative superior performance by the VECM.

	VECM (level)	VAR (ratio)	Random Walk (ratio)	Forward Premium Regression (ratio)	Forward Rate (ratio)
Root-mean-square error (RMSE)					
4-week horizon	0.0258	1.005	0.968	0.976	0.968
13-week horizon	0.0573	0.780	1.057	1.020	1.057
26-week horizon	0.0834	0.206	1.205	1.063	1.192
52-week horizon	0.1079	0.003	1.257	1.088	1.207
Mean absolute error (MAE)					
4-week horizon	0.0207	1.023	1.000	0.976	1.001
13-week horizon	0.045	0.763	1.075	1.019	1.072
26-week horizon	0.0676	0.210	1.300	1.030	1.273
52-week horizon	0.0888	0.003	1.448	1.066	1.381

3.7 Conclusion

Contrary to the findings in the original article we are unable to produce forecasts that have RMSE and MAE statistics that are significantly lower than that of a naïve random walk. An important underlying reason for the observed differences appears to stem from the data that are used since the statistical results begin to deviate from the outset of the empirical implementation. Alternatively, an error was made in the implementation of the relevant statistical algorithms in which case we believe that this error was made at the beginning of the

empirical study. Any possible subsequent errors are impossible to discern from the reported information. Further, there appears to be quite a large amount of uncertainty with respect to the selected models and the other statistics that were calculated. Importantly, when testing the empirical validity of the proposed model CT do not indicate which data series they use for the spot rate. The Harris Bank database reports three different series: the closing bid price for the week, the spot high during the week and the spot low during the week. As stated we performed the analysis using all three series as well as the mid-prices and find that we are unable to significantly outperform the random walk at any horizon for any of these series. We view this as evidence that there is significantly less empirical support for the framework than is claimed.

4

An Expedient Fitting Approach in Multiple Threshold TAR Time Series Modeling

4.1. Introduction

The hegemony of linear cointegration techniques has inexorably led to the lack of empirical substantiation of many interesting economic hypotheses. Unstinting efforts have consequently concentrated on nonlinear extensions of extant techniques. An example of this is the development of threshold cointegration, ascribed to Balke and Fomby (1997), and the statistical tests implicit to the technique which rely heavily on the estimation of TAR models, or variants thereof (see Enders and Granger (1998), Gonzalez and Gonzalo (1998), Berben and van Dijk (1999), amongst others, for unit root tests specifying a (SE)TAR model under the alternative). Computational tractability is a particularly pertinent issue when estimating models of this type, however, especially in the context of Monte Carlo simulations, possibly augmented with the need for bootstrapping techniques. The reason for this is the inapplicability of standard maximum-likelihood estimation techniques, which, consequently,

implies the need for a Grid Search (GS) over a feasible region of a continuous parameter space.

In this chapter we advocate a generalization of the fitting approach proposed by Coakley, Fuertes and Perez (2003) (henceforth CFP), who tackle Sequential Conditional Least Squares (SCLS) problems by using Givens transformations to update QR factorizations and, based on their proof that the residual sum of squares function is both continuous and rational over non-overlapping threshold intervals, estimate the location of the threshold by subjecting a rational interpolation function to the standard optimality condition. The discussion, initialized by introducing relevant definitions and nomenclature, will begin by expositing the fitting approach advocated by CFP in §4.3. An algorithm for the construction of an appropriate higher-dimensional threshold grid, when estimating SETAR models, including a sorting algorithm which allows for the simple implementation of updating algorithms is presented and discussed in §4.4. Utilizing this grid to construct a set of non-overlapping threshold hypercubes is then detailed in §4.5. Based on these constructs we develop a novel, numerical, rational interpolation-based optimization routine after discussing the applicability of QR factorizations and Givens transformations to the multiple-threshold SETAR model context in §4.6 Ensuingly, §4.7 entails a presentation and discussion of the proposed fitting approach, while Monte Carlo simulations are utilized, in §4.8, to assess the performance of the developed approach relative to a set of viable alternatives, including the conventionally adopted SCLS routine.

Our results suggest that the proposed approach is superior to all alternatives in terms of estimation time for five- and six-regime SETAR models, while a composite fitting approach, SCLS in combination with the proposed optimization routine, alluded to previously, is more adept at expediently fitting three- and four-regime SETAR models. The results are astonishing: conservative estimates indicate that the proposed approach is 570 and 14 000 times faster than the conventionally adopted approach when fitting five- and six-regime models, respectively. The composite approach is 2.3 and 32.6 times faster in fitting three- and four-regime models, respectively. These computation time estimates are attained while achieving similar degrees of model fit in terms of a set of bias measures when utilizing SCLS[18].

[18] Importantly, these results were attained while minimizing cross-program variation; hence rendering the results (approximately) free of differences in programming efficiency.

4.2. Definitions and Nomenclature

A time series z_t is said to be a *k*-regime TAR process if it follows the model:

$$z_t = \sum_{j=1}^{k} \left[\phi_0^{(j)} + \phi_1^{(j)} z_{t-1} + \ldots + \phi_{p_j}^{(j)} z_{t-p^{(j)}}\right] I_t\left(\theta^{(j-1)} \le v_{t-d} < \theta^{(j)}\right) + \varepsilon_t \tag{4.1}$$

where ε_t is assumed to be a martingale difference sequence with respect to the past history of z_t; j, d and $p^{(j)}$ are positive integers; $I_t(A)$ is an indicator function such that it equals unity when the event A occurs and is zero otherwise, $\phi_i^{(j)}$ denotes the autoregressive coefficient of the *ith* lag in regime *j*; while v_{t-d} denotes the delay variable. If $v_{t-d} \equiv z_{t-d}$ then (4.1) is typically referred to as a self-exciting TAR (SETAR) model. Essentially, the model partitions the one-dimensional Euclidean space into *k* linear AR regimes based on the value of the delay parameter relative to the $(k+1)$-dimensional threshold vector $\boldsymbol{\theta} = \left[\theta^{(0)}, \theta^{(1)}, \ldots, \theta^{(k)}\right]$, where $\theta^{(0)} = -\infty$ and $\theta^{(k)} = \infty$.

For need of abutting the discussion in succeeding sections two alternative SETAR models will be utilized throughout; these cases are easily generalized. Consider the following first-difference reparameterizations of (4.1):

$$\begin{aligned}\Delta z_t = &\left[\phi_0^{(1)} + \sum_{i=1}^{p^{(1)}} \phi_i^{(1)} z_{t-i}\right] I_t\left(z_{t-1} < \theta^{(1)}\right) + \left[\phi_0^{(2)} + \sum_{i=1}^{p^{(2)}} \phi_i^{(2)} z_{t-i}\right] I_t\left(\theta^{(1)} \le z_{t-1} < \theta^{(2)}\right) + \\ &\left[\phi_0^{(3)} + \sum_{i=1}^{p^{(3)}} \phi_i^{(3)} z_{t-i}\right] I_t\left(z_{t-1} \ge \theta^{(2)}\right) + \varepsilon_t\end{aligned} \tag{4.2}$$

$$\begin{aligned}\Delta z_t = &\left[\sum_{i=1}^{p^{(1)}} \phi_i^{(1)} (z_{t-i} - \theta^{(1)})\right] I_t\left(z_{t-1} < \theta^{(1)}\right) + \left[\phi_0^{(j)} + \sum_{i=1}^{p^{(2)}} \phi_i^{(2)} z_{t-i}\right] I_t\left(\theta^{(1)} \le z_{t-1} < \theta^{(2)}\right) \\ &\left[\sum_{i=1}^{p^{(3)}} \phi_i^{(3)} (z_{t-i} - \theta^{(2)})\right] I_t\left(z_{t-1} \ge \theta^{(2)}\right) + \varepsilon_t\end{aligned} \tag{4.3}$$

where, for convenience, it is assumed that $\varepsilon_t \sim IIN\left(0,\sigma_\varepsilon^2\right)$ while Δ denotes the first-difference operator i.e. $\Delta z_t = z_t - z_{t-1}$. Notationally, the remainder may be inferred from (4.1). Models (4.2) and (4.3) are three-regime SETAR models typically referred to as Equilibrium-TAR and Band-TAR models[19], respectively, since being coined as such by Balke and Fomby (1997)[20]. Subjecting (4.1) to the requirement that the piecewise linear autoregressive function be continuous everywhere, results in an important TAR sub-class, the continuous TAR or CTAR model, introduced into the literature by Chan and Tsay (1998) [21]. Juxtaposing the notions demanded by the proposed generalization with those of the CFP procedure justifies the inclusion of added notational complexity. Consequently, we introduce important definitions and discuss notational convention at this point. The kernel of the discussion is the concepts and constructs related to the continuous threshold space around which the procedure convolutes.

Fitting SETAR models entails searching for the most likely location of the threshold vector, $\boldsymbol{\theta}$, in the continuous threshold space Θ [22]. Hence, we embed a grid of discrete points in the form of a $\left(G\times(k-1)\right)$ structured matrix, denoted $\boldsymbol{\Theta}$, in which the *gth* row vector defines a candidate threshold vector, denoted $\boldsymbol{\theta}_g$, the elements of which will frequently be referred to as $\theta_g^{(i)}$ for $i=1,2,\ldots,k-1$ (also $\boldsymbol{\theta}_g$ is an ordered row vector such that $\theta_g^{(1)} < \theta_g^{(2)} < \ldots < \theta_g^{(k-1)}$). Alternatively, each $\boldsymbol{\theta}$ may be interpreted as a threshold location in Θ relative to a $(k-1)$-axis Cartesian coordinate system, the axes of which will be referred to as threshold axes.

The generalization that we propose requires the interpolation of a rational function, consequently it necessitates a series of discrete points, $\theta_{\mathbf{H}_g,q}^{(i)}$ for $q \in \{1,2,\ldots\max(q)\}$, lying on a line, parallel to the *ith* threshold axis. The proposed optimization routine will make use of a

[19] We refer readers to the Balke and Fomby (1997) article on threshold cointegration for a discussion on the dynamics of models of this type. Note that our presentation of these models deviates from theirs since we have generalized the models to allow for asymmetric dispersion of the thresholds around the unconditional mean of the process and differing autoregressive coefficients in the outer-regimes, implying differing speeds in equilibrium reversion (under the assumption that the process is globally stationary).

[20] §2.3.2 details the conditions that are required to ensure stationarity in models of this type.

[21] Instituting the restrictions $p^{(j)}=1$ for all j; $\phi_0^{(2)}=0$; and $\phi_1^{(2)}=1$ results in three-regime SETAR models that allow for unit root behavior in the inner-regime and autoregressive behavior in the outer-regimes. It is easily shown that restricting (4.3), in this manner, will result in a model that meets the "continuous everywhere" requirement implying that it belongs to the CTAR sub-class.

[22] Effectively, the threshold parameter space is restricted by the distribution of the delay variable.

directional component whereby $\bar{\theta}_g^{(i)}$ indicates that the domain of the interpolated function is in a direction parallel to the *ith* threshold axis.

Generally, superscripting a variable with (j), for $j = 1,2,\ldots,k$, indicates to which regime it refers, while subscripting a variable with g indicates that the active candidate threshold vector, $\boldsymbol{\theta}_g$, or, more precisely, the threshold hypercube $\mathbf{H}_g$, has been activated[23]. Hence, if the sample size and the effective sample size are denoted by N and N^*, respectively, then $N^* = N - \max\left(d, p^{(j)}\right)$ for $j = 1,2,\ldots,k-1$, and the effective sample size in regime j may be denoted by $N_g^{*(j)}$, for some $\boldsymbol{\theta}_g$, where, logically, $N^* = \sum_{j=1}^{k} N_g^{*(j)}$ for any g.

If the sub-model present in regime j is an explicit function of the threshold vector *and* if the regressor matrix may be represented as a first-degree polynomial matrix[24], denoted $\mathbf{F}_g^{(j)}$, then it may be decomposed as $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}$ where

$$\mathbf{F0}_g^{(j)} = \begin{bmatrix} z_{1t-1}^{(j)} & z_{1t-2}^{(j)} & \cdots & z_{1t-L}^{(j)} \\ z_{2t-1}^{(j)} & & & \vdots \\ \vdots & & & \\ z_{N_g^{*(j)}t-1}^{(j)} & \cdots & & z_{N_g^{*(j)}t-L}^{(j)} \end{bmatrix}; \tag{4.4}$$

$z_{ab}^{(j)}$ denotes the regressor value for case[25] a at lag b in regime j; L is the maximum feasible dimensionality of the lag space; and the $\left(N_g^{*(j)} \times L\right)$ matrix $\mathbf{U}_g^{(j)} = [c,c,\ldots,c]'\left[\theta_g^{(i)}, \theta_g^{(i)}, \ldots, \theta_g^{(i)}\right]$ where c is a scalar. Alternatively, if $\mathbf{F}_g^{(j)}$ is not an explicit function of the threshold vector then $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)}$. For purposes of intelligibility we will refer to $\mathbf{F0}_g^{(j)}$ as the regressor matrix of regime j, given $\boldsymbol{\theta}_g$, when it is not a function of the threshold vector.

Finally, it is convenient to denote a n-vector of ones by $\boldsymbol{\iota}_n$; an identity matrix of order n by $\mathbf{I}_n$; and a floor function by $\lfloor \cdot \rfloor$ denoting the integer part operator[26]. The next section delineates our point of departure: the fitting approach advocated by CFP.

---

[23] The exact definition of the threshold hypercube $\mathbf{H}_g$ is presented in §4.5.

[24] This is the case for the outer-regimes of model (4.3), for example.

[25] In the sense of Tsay (1989).

[26] A good reference book for the matrix manipulations of this chapter is Golub and van Loan (1990).

4.3. Fitting a Symmetric Three-Regime Band-TAR Process

Typically, (SE)TAR models are estimated by computing parameter estimates sequentially over a set of plausible threshold vectors, where the elements of each vector have been restricted to being members of the set of order statistics of the delay variable, and selecting those parameter estimates that correspond with a minimization of the residual variance. SCLS estimation will correspond with maximization of the likelihood function if Δz_t is not an explicit function of $\boldsymbol{\theta}$, as is the case in (4.2) above, since varying the elements of $\boldsymbol{\theta}$ between the order statistics of the delay variable will not change how each case[27] is partitioned into the various regimes. Model (4.3), on the other hand, where Δz_t is an explicit function of $\boldsymbol{\theta}$, requires a full GS, with a sufficiently small stepsize, to maximize the likelihood function.

To keep the time needed for the estimation of these models within tractable limits CFP have developed an efficient estimation procedure based on computing values for the residual sum of squares (RSS) function using Givens transformations of QR factorizations and inputting these into a rational interpolation algorithm, which is subsequently subjected to the standard optimality condition.

Below we briefly introduce and discuss this approach, note that the approach is specific to the general form of a *symmetric* three-regime Band-SETAR specification, i.e. $\theta = \theta^{(2)} = -\theta^{(1)}$, where $\theta > 0$ is an identifying restriction, and we set $\phi_i^{(1)} = \phi_i^{(3)}$ for all $i \le p^{(1,3)}$ with $p^{(1,3)} = p^{(1)} = p^{(3)}$. The authors identify the following steps to identify and estimate the Band-SETAR parameters:

1) Let D and L denote the maximum feasible values for the threshold delay and AR lag orders, respectively. Set the restriction on the minimum number of observations that must be present in each regime m such that $m = N\kappa$ where, typically, $\kappa = 0.15$. See §4.4 for the implications of the requirement κ on the construction of an appropriate grid.
2) For each $d \in \{1,2,\ldots,D\}$ repeat the following:
   a) Arrange the data in ordered autoregression form, following the suggestion of Tsay (1989). Let $\mathbf{F}_g^{(2)}$ and $\mathbf{F}_g^{(1,3)}$ define the $\left(N_g^{*(2)} \times (L+1)\right)$ inner and $\left(\left(N_g^{*(1)} + N_g^{*(3)}\right) \times L\right)$ outer regressor matrices, respectively, defined as a simple partitioning of the arranged

[27] In the sense of Tsay (1989).

data and g denotes the *gth* candidate threshold. Note that the outer-regime regressor matrix $\mathbf{F}_g^{(1,3)} = \mathbf{F0}_g^{(1,3)} + \mathbf{U}_g^{(1,3)}$ where the matrix $\mathbf{A}_g^{(i,j)}$ is the result of vertically concatenating matrix $\mathbf{A}_g^{(i)}$ with $\mathbf{A}_g^{(j)}$ and $\mathbf{U}_g^{(1,3)} = (1,\ldots,1,-1,\ldots,-1)'(\theta,\ldots,\theta,\theta,\ldots,\theta)$. The authors refer to $\mathbf{F0}_g^{(j)}$ and $\mathbf{U}_g^{(j)}$ as basis and correction components, respectively.

b) Let $\mathbf{H}_g = \lfloor \theta_g, \theta_{g+1} \rfloor$, where $\theta_g = z_{(g)}$ i.e. θ_g is defined as the *gth* smallest order statistic of the delay variable, then, the continuous threshold space is defined as $\Theta = \{\cup_g \mathbf{H}_g,\ g = \tau_0, \tau_0 + 1, \ldots, \tau_1 - 1\}$ for $\tau_0 = \lfloor N\pi_0 \rfloor$ and $\tau_1 = \lfloor N(1-\pi_1) \rfloor$ where $\{\pi_0, \pi_1\}$ are trimming parameters. Hence, the threshold space is defined as a union of continuous non-overlapping threshold intervals.

c) Use $\mathbf{H}_{\tau_0}$ as the initial threshold interval, whereby we allow $\theta = \theta_{\tau_0}$ to be the first threshold candidate, and perform the following operations for the outer-regimes over each threshold interval, progressing sequentially through each threshold interval in Θ.

i) Filter out the outer-regime cases, given θ, to form the augmented basis matrix $\left[\mathbf{F0}_g^{(1,3)} \mid \Delta\mathbf{z}_g^{(1,3)}\right]$ for orders $p^{(1,3)} = L$ and compute its QR factorization.

ii) Choose 7 arbitrary thresholds $\theta_{\mathbf{H}_g,q} \in \mathbf{H}_g$ for $q = 1,2,\ldots,7$.

iii) Factorize the augmented regressor matrix $\left[\mathbf{F}_g^{(1,3)} \mid \Delta\mathbf{z}_g^{(1,3)}\right]$ for $q = 1,2,\ldots,7$ by means of rank-one Givens updates of the QR factorization of $\left[\mathbf{F0}_g^{(1,3)} \mid \Delta\mathbf{z}_g^{(1,3)}\right]$.

iv) Compute $RSS_g^{(1,3)}\left(\theta_{\mathbf{H}_g,q}, d, p^{(1,3)}\right)$ for $q = 1,2,\ldots,7$.

v) Identify $\psi_g^{4,2(1,3)}(\theta)$, a rational function of degree type $(4,2)$, via rational interpolation with support points $\left(\theta_{\mathbf{H}_g,q}, RSS_g^{(1,3)}\left(\theta_{\mathbf{H}_g,q}, d, p^{(1,3)}\right)\right)$ for $q = 1,2,\ldots,7$.

vi) Find θ_g^*, the value that minimizes $\psi_g^{4,2(1,3)}(\theta)$ in $\mathbf{H}_g$ using the necessary optimality condition. Compute the associated $RSS_g^{(1,3)}\left(\theta_g^*, d, p^{(1,3)}\right)$ and $AIC_g^{(1,3)}\left(\theta_g^*, d, p^{(1,3)}\right)$.

vii) Set $p^{(1,3)} = p^{(1,3)} - 1$. Apply Givens rotations to the QR factorization of $\left[\mathbf{F}_g^{(1,3)} \mid \Delta\mathbf{z}_g^{(1,3)}\right]$ to re-factor the new basis matrix.

viii) Repeat steps (ii) – (vii) until $p^{(1,3)} = 1$ is completed.

ix) Find $\tilde{p}^{(1,3)}$ (and associated optimal θ_g^*), the best-fit lag order for the active threshold interval as the value that minimizes the $AIC_g^{(1,3)}\left(\theta_g^*, d, p^{(1,3)}\right)$ sequence:

$$\tilde{p}^{(1,3)} = \underset{1 \le p^{(1,3)} \le L}{\arg\min}\, AIC_g^{(1,3)}\left(\theta_g^*, d, p^{(1,3)}\right)$$

and compute the minimal $AIC_g^{(1,3)}\left(\theta_g^*, d, \tilde{p}^{(1,3)}\right)$.

x) Move to the next threshold interval and re-factor the new basis $\left[\mathbf{F0}_{g+1}^{(1,3)} \mid \Delta\mathbf{z}_{g+1}^{(1,3)}\right]$ for $p^{(1,3)} = L$ by applying Givens rotations to the QR decomposition of the previous interval's basis component.

d) Perform the following operation for the inner regime:

i) Use $\mathbf{H}_{r_1-1}$, the last interval used in the outer-regime iterations, to filter out the inner-regime cases to form the augmented regressor matrix $\left[\mathbf{F0}_g^{(2)} \mid \Delta\mathbf{z}_g^{(2)}\right]$ for order $p^{(2)} = L$. Compute its QR factorization.

ii) Use the R factor of $\left[\mathbf{F0}_g^{(2)} \mid \Delta\mathbf{z}_g^{(2)}\right]$ to compute $RSS_g^{(2)}\left(\theta_g, d, p^{(2)}\right)$ and the associated $AIC_g^{(2)}\left(\theta_g, d, p^{(2)}\right) = N_g^{*(2)} \ln\left(RSS_g^{(2)}\left(\theta_g, d, p^{(2)}\right)/N_g^{*(2)}\right) + 2\left(p^{(2)} + 1\right)$.

iii) Set $p^{(2)} = p^{(2)} - 1$ and apply Givens rotations to the R factor of $\left[\mathbf{F0}_g^{(2)} \mid \Delta\mathbf{z}_g^{(2)}\right]$ to re-factor the new basis matrix.

iv) Repeat steps (ii) and (iii) until $p^{(2)} = 1$.

v) Determine $\tilde{p}^{(2)}$, the best-fit order for the active threshold interval as the value that minimizes the $AIC_g^{(2)}\left(\theta_g, d, p^{(2)}\right)$ sequence:

$$\tilde{p}^{(2)} = \underset{1 \le p^{(2)} \le L}{\arg\min}\, AIC_g^{(2)}\left(\theta_g, d, p^{(2)}\right)$$

and compute the minimal $AIC_g^{(2)}\left(\theta_g, d, \tilde{p}^{(2)}\right)$.

vi) Move to the next (contiguous) threshold interval in Θ. Obtain the R factor of the new $\left[\mathbf{F0}_g^{(2)} \mid \Delta\mathbf{z}_g^{(2)}\right]$ for $p^{(2)} = L$ by applying Givens rotations to the R factor of the same matrix defined over the previous interval.

vii) Repeat steps (ii) – (vi), allowing each threshold in Θ to become sequentially activated.

e) For each threshold interval compute an overall AIC where

$$AIC_g^{TAR}\left(\theta_g^*, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right) = AIC_g^{(1,3)}\left(\theta_g^*, d, \tilde{p}^{(1,3)}\right) + AIC_g^{(2)}\left(\theta_g, d, \tilde{p}^{(2)}\right)$$

Minimize $AIC_g^{TAR}\left(\theta_g^*, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right)$ conditional on d, across intervals to find:

$$\left(\tilde{\theta}, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right) = \arg\min_{1 \le g \le G} AIC_g^{TAR}\left(\theta_g^*, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right)$$

and calculate the associated minimal normalized AIC ($N\,AIC$) as

$$N\ AIC^{TAR}\left(\tilde{\theta}, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right) = AIC^{TAR}\left(\tilde{\theta}, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right) / \left(N^*\right)$$

where $AIC^{TAR}\left(\tilde{\theta}, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right) = \min_{1 \le g \le G} AIC_g^{TAR}\left(\theta_g^*, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right)$, following Tong and Lim (1980).

3) The LS estimates $\hat{\theta}$, $\hat{d}$, $\hat{p}^{(1,3)}$ and $\hat{p}^{(2)}$ are obtained by minimizing $N\,AIC$:

$$\left(\hat{\theta}, \hat{d}, \hat{p}^{(1,3)}, \hat{p}^{(2)}\right) = \arg\min_{1 \le d \le D} N\ AIC^{TAR}\left(\tilde{\theta}, d, \tilde{p}^{(1,3)}, \tilde{p}^{(2)}\right)$$

4) Finally, estimate the autoregressive coefficients of the Band-TAR model, conditional on $\left(\hat{\theta}, \hat{d}, \hat{p}^{(1,3)}, \hat{p}^{(2)}\right)$ by ordinary LS employing all available data.

The authors assess the contribution of the approach, with respect to computational expense and estimation bias, relative to a GS using QR factorizations and Givens updates (without interpolating the RSS function); a GS using QR factorizations (without Givens updates nor interpolation of the RSS function); and a simple GS using SCLS. They find that the proposed approach mitigates small sample bias at the expense of taking longer to estimate relative to the GS using QR factorizations and Givens updates; which was the most expedient of the different approaches. Hence, an important finding of their study is that using QR factorizations may be useful when the focal point of interest is computational expense.

Both theoretically and empirically it may be desirable that the procedure have the capacity to allow for a multidimensional threshold vector, as is the case when the thresholds of a three-regime SETAR process are asymmetrically dispersed around its unconditional mean. In the

sections that follow we aim to generalize, the previously delineated, procedure so that it may be utilized for parameter estimation of multiple-threshold, possibly asymmetric, SETAR processes. We begin by developing an appropriately constrained multidimensional grid which will form the foundation of the proposed generalization.

4.4. Constructing an Appropriate Grid

In this section we design a computationally efficient procedure for constructing a set of candidate threshold vectors, of any desired dimensionality, which are in compliance with requirements that are relevant when estimating the parameters of a (SE)TAR model. Letting Π denote a set whose closure lies in $(0,1)$, then, based on asymptotic convergence arguments, Andrews (1993) shows that if no information is available regarding the location of the thresholds it is desirable to bound Π away from 0 and 1, where Π is the set of possible change points defined as a fraction of N^*. The trimming parameters $\pi_0, \pi_1 (\in \Pi)$ restrict the set of candidate threshold vector elements such that $z_{(\lfloor N^* \pi_0 \rfloor)} < \theta_g^{(i)} < z_{(\lfloor (1-\pi_1) N^* \rfloor)}$ for all i and g. Andrews (1993) consequently suggests setting $\pi_0 = \pi_1 = 0.15$. Further, if we define $\boldsymbol{\Phi}^{(j)} = \left(\phi_0^{(j)}, \phi_1^{(j)}, \ldots, \phi_{p^{(j)}}^{(j)}\right)'$ as the coefficient vector of regime j, then it is apparent that to achieve an acceptable degree of accuracy in computing $\hat{\boldsymbol{\Phi}}^{(j)}$ for $j = 1,2,\ldots,k$, the estimates of $\boldsymbol{\Phi}^{(j)}$, a minimum proportion, say κ, of the N^* observations must be present in each regime. Let $m = \lfloor N^* \kappa \rfloor$, then there must be at least m observations in each regime. For purposes of clarity we briefly state these two requirements:

1. The trimming parameters π_0 and π_1 define a restriction on the norm of $\boldsymbol{\theta}_g$, if we define $\tau_0 = \lfloor N^* \pi_0 \rfloor$ and $\tau_1 = \lfloor (1-\pi_1) N^* \rfloor$ then this implies $z_{(\tau_0)} < \theta_g^{(i)} < z_{(\tau_1)}$ where $\theta_g^{(i)} \in \boldsymbol{\theta}_g$ for $i = 1,2,\ldots,k-1$ and all g.
2. At least m observations must be present in each of the k regimes.

Clearly, for the outer-regimes of a TAR model the binding constraint holds i.e. requirement 1 holds unless $m > \tau_0$ or $m > N^* - \tau_1$ in which case the second requirement is binding. A grid meeting these requirements will form an embedded subset of points in the continuous threshold

space Θ, over which the GS is ultimately performed. Heuristically, we will combine the candidate thresholds, $\theta_g^{(i)}$, into candidate threshold vectors, $\boldsymbol{\theta}_g$, meeting the stipulated requirements, which will ultimately constitute the rows of $\boldsymbol{\Theta}$.

4.4.1. Constructing $\boldsymbol{\Theta}$

The proposed combinatorial procedure will combine 2-dimensional subsets of candidate thresholds until the desired dimensionality of the candidate threshold vector is achieved. The following steps can be followed to construct the matrix $\boldsymbol{\Theta}$:

1. If $k = 2$
   a. If $\kappa > \pi_i$ for $i = 0$ or 1 set $\pi_i = \kappa$. The set of order statistics, ranging from $z_{(\tau_0)}$ through $z_{(\tau_1)}$, of the delay variable define the gridpoints of the grid, after having removed any non-unique elements.
2. If $k > 2$
   a. Let $\overline{N} = u - l - m$, where $l = \max(\tau_0, m)$ and $u = \min(\tau_1, N^* - m)$, then we use the following subset of the order statistics $\{z_{(l)}, z_{(l+1)}, \ldots, z_{(u)}\}$ to construct a $\overline{N} \times 2$ matrix upon which the approach builds. It is useful to define

$$\mathbf{o}_{[i,j]}^{\mathbf{z}} = \left[z_{(i)}, z_{(i+1)}, \ldots, z_{(j)}\right]',$$

   where $\min(i) = l$, $\max(j) = u$ and $i < j$, as a (sub-) vector of this restricted set of order statistics. Fix the values of u and l, then the augmented matrix which we wish to construct is given as $\left[\mathbf{o}_{[l,u-m]}^{\mathbf{z}} \mid \mathbf{o}_{[l+m,u]}^{\mathbf{z}}\right]$. Each row in this matrix defines two candidate thresholds which are separated by exactly m observations.
      - Let $m = m + 1$ and define a series of matrices in an analogous manner, until $\overline{N} = 1$.
   b. Vertically concatenate these matrices into a matrix which we will refer to as the pre-grid in what is to follow. Note that each row in the pre-grid is defined

by two candidate thresholds such that both of the aforementioned requirements are complied with.

c. If $k = 3$

then $\boldsymbol{\Theta}$ is the pre-grid.

If $k > 3$,

then duplicate the pre-grid matrix labeling the resultant matrices G^0 and G^1, subsequently consider each row in G^0 sequentially and match it with all rows in G^1, for which it is true that the second element of the former is equal to the first element of the latter.

d. Label the resultant matrix G^0 and repeat step c until the desired dimensionality of the candidate threshold vectors is achieved.

The resultant matrix, $\boldsymbol{\Theta}$, has a $(k-1)$-dimensional candidate threshold vector in every row and contains the full set of possible candidate threshold vectors which comply with the stated requirements. The candidate threshold vectors of the constrained grid, thus constructed, will, however, need to be reordered to facilitate the harmonious implementation of updating algorithms. Details on these updating algorithms and their applicability to the case at hand are discussed in §4.6.

4.4.2. Facilitating the Use of Updating Algorithms: Resorting $\boldsymbol{\Theta}$

Let $\theta_{(i)} = z_{(i)}$ for $i \in (l,\ldots,u)$ then $\theta_{(i)}$ implicitly denotes the *ith* order statistic of the delay variable. Considering this set of plausible candidate thresholds it is apparent that each $\theta_{(i)}$ is associated with the index number i based on its position in the set of order statistics of the delay variable. The underlying scheme, of the algorithm presented below, is to reorder the row vectors of $\boldsymbol{\Theta}$ so that when moving from $\boldsymbol{\theta}_g$ to $\boldsymbol{\theta}_{g+1}$ only two contiguous regimes are affected by a repartitioning of the cases.

The proposed sorting algorithm is presented below:

1) Sort the row vectors of $\boldsymbol{\Theta}$ by sorting $\boldsymbol{\theta}^{(i)}$ in ascending order for $i = 1, 2, \ldots, k-1$ sequentially, where $\boldsymbol{\theta}^{(i)}$ denotes the *ith* column of $\boldsymbol{\Theta}$.

2) Let $\mathbf{I}^{\Theta}$ denote a matrix, equal in order to that of $\boldsymbol{\Theta}$, whereby each element $\theta_{(i)} \in \boldsymbol{\Theta}$ has been substituted by its (unique) index number[28] *i*. Further, let $r = \max(i) - (k-2)m$ and T_j be a Triangular number i.e. $T_j = \frac{1}{2} j(j+1)$. Subsequently, generate the following structured matrix consisting of *k-2* column vectors where the recurrence relation is most easily derived from the matrix presented below:

$$\begin{bmatrix} \overline{\overline{T}}_r = \sum_{j=1}^{r} \overline{T}_j & \overline{T}_r = \sum_{j=1}^{r} T_j & T_r & r \\ \overline{\overline{T}}_{r-1} = \sum_{j=1}^{r-1} \overline{T}_j & \overline{T}_{r-1} = \sum_{j=1}^{r-1} T_j & T_{r-1} & r-1 \\ \overline{\overline{T}}_{r-2} = \sum_{j=1}^{r-2} \overline{T}_j & \overline{T}_{r-2} = \sum_{j=1}^{r-2} T_j & T_{r-2} & r-2 \\ \vdots & \vdots & \vdots & \vdots \\ \overline{\overline{T}}_2 = \sum_{j=1}^{2} \overline{T}_j & \overline{T}_2 = \sum_{j=1}^{2} T_j & T_2 & 2 \\ \overline{\overline{T}}_1 = \sum_{j=1}^{1} \overline{T}_j & \overline{T}_1 = \sum_{j=1}^{1} T_j & T_1 & 1 \end{bmatrix}$$

As is evident this matrix, denoted $\mathbf{T}$, is easily generated once *r* is known.

3) Construct a matrix $\mathbf{C}$, equal in order to $\mathbf{T}$, such that $\mathbf{C}_{i,j} = \sum_{l=i}^{r} \mathbf{T}_{l,j}$ where $\mathbf{T}_{i,j}$ and $\mathbf{C}_{i,j}$ denote the *ith* element of the *jth* column in the matrices $\mathbf{T}$ and $\mathbf{C}$, respectively. Finally, redefine $\mathbf{C}$ such that $\mathbf{C} = \left[\mathbf{t}_{k-2} \mid \mathbf{C}\right]$.

4) Each element of $\mathbf{C}$ partitions the matrix $\mathbf{I}^{\Theta}$ into submatrices defined as $\mathbf{A}_{i,j} = \left[\mathbf{I}^{\Theta}_{C_{i,j}, C_{i+1,j}}\right]$ where $C_{i,j}, C_{i+1,j}$ define the row indices of $\mathbf{I}^{\Theta}$ that form the first and last row vector of submatrix $\mathbf{A}_{i,j}$, respectively. Thus, $\mathbf{A}_{i,j}$ is a $(C_{i+1,j} - C_{i,j}) \times (k-1)$ submatrix of $\mathbf{I}^{\Theta}$.

5) Starting with $j = 1$ consider each $i = 1,2,\ldots,r$ in order, and reverse the order of the row vectors of submatrix $\mathbf{A}_{i,j}$ if $i \bmod(2) = 0$, redefining the resultant matrix as $\mathbf{I}^{\Theta}$.

6) Reapply step 5) for $j = 2,3,\ldots,k-2$ in order.

[28] Logically, one could also generate $\mathbf{I}^{\Theta}$ and subject it to the sorting algorithm described in 1).

7) Define the vector $\boldsymbol{\Delta}$, where the *ith* element, Δ_i, is defined as $\Delta_i = \iota'_{k-1}\left(\left|\mathbf{I}_i^{\Theta} - \mathbf{I}_{i-1}^{\Theta}\right|\right)$ unless $i = 1$ in which case $\Delta_1 = 1$.

i) If $\Delta_i = 1$, the location of the active candidate threshold vector will change to either a previous/subsequent candidate threshold vector by moving from the threshold vector associated with $\mathbf{I}_{i-1}^{\Theta}$ to that associated with $\mathbf{I}_i^{\Theta}$. From the perspective of the updating algorithms this case is most suitable since only two regimes will be affected by a repartitioning of the cases.

ii) If $\Delta_i \neq 1$, new row vectors will need to be inserted into $\mathbf{I}^{\Theta}$ until the requirement that $\Delta_i = 1$ is met. This is done by inserting a suitably constructed matrix between $\mathbf{I}_i^{\Theta}$ and $\mathbf{I}_{i-1}^{\Theta}$. Initialize a $\left(\left(\iota'_{k-1}\left(\left|\mathbf{I}_i^{\Theta} - \mathbf{I}_{i-1}^{\Theta}\right|\right) - 1\right) \times k\right)$ matrix of zeros and denote it $\mathbf{I}^0$; let $\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}$ denote the *jth* difference in the row vector $\mathbf{I}_i^{\Theta} - \mathbf{I}_{i-1}^{\Theta}$; and $j = k - 1$ then

(a) If $\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta} > 1$: substitute the first $\left|\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}\right|$ zeros of the *jth* column of $\mathbf{I}^0$ with $\iota_{\left|\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}\right|}$.

(b) If $\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta} < 1$: substitute the first $\left|\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}\right|$ zeros of the *jth* column of $\mathbf{I}^0$ with $-\iota_{\left|\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}\right|}$.

(c) Let $j = j - 1$ and repeat steps a) and b) until $\left|\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}\right| = 1$, by substituting a vector of ones into the *jth* column of the next $\left|\mathbf{I}_{i,j}^{\Theta} - \mathbf{I}_{i-1,j}^{\Theta}\right|$ rows of $\mathbf{I}^0$, and so forth.

iii) Let the $\left(\left(\iota'_{k-1}\left(\left|\mathbf{I}_i^{\Theta} - \mathbf{I}_{i-1}^{\Theta}\right|\right) - 1\right) \times k\right)$ matrix that we intend to insert between $\mathbf{I}_i^{\Theta}$ and $\mathbf{I}_{i-1}^{\Theta}$ be denoted $\bar{\mathbf{I}}$, then, if $\bar{\mathbf{I}}_b$ denotes the *bth* row of this matrix, $\bar{\mathbf{I}}_1 = \mathbf{I}_{i-1}^{\Theta} + \mathbf{I}_1^0$, $\bar{\mathbf{I}}_2 = \mathbf{I}_1^0 + \mathbf{I}_2^0$, $\bar{\mathbf{I}}_3 = \mathbf{I}_2^0 + \mathbf{I}_3^0$, and so forth.

iv) Insert $\bar{\mathbf{I}}$ between $\mathbf{I}_{i-1}^{\Theta}$ and $\mathbf{I}_i^{\Theta}$, after which step 7) is repeated for all *i*.

8) $\mathbf{I}^{\Theta}$ may be retransformed into an updated grid, by associating $\theta_{(i)}$ with each of its corresponding index values defined as elements in $\mathbf{I}^{\Theta}$. Henceforth, $\boldsymbol{\Theta}$ will denote this retransformed updated grid.

As mentioned previously, if the functions representing the processes in each regime are independent of the threshold vector the grid that results when mapping $\mathbf{I}^{\Theta}$ back to the threshold space will suffice for purposes of parameter estimation. If this is not the case, however, the interior of a set of hypercubes, the vertices of which are defined by 2^i candidate threshold vectors, where $i = 0,\ldots,k-1$, will also need to be searched. To this end, a multidimensional interpolation-based optimization routine will be used to locate this point. We begin by utilizing $\mathbf{I}^{\Theta}$ to construct the relevant hypercubes, which, by construction, will comply with the stated requirements.

4.5. Multidimensional Non-Overlapping Threshold Hypercubes

It is easily shown that the "locally continuous" property of the rational $RSS_g^{(1,2,\ldots,k)}(\mathbf{H}_g, d, \mathbf{p}_i)$ function, in the single threshold case, also applies in the context of higher-dimensional threshold vectors[29], implying the applicability of a suitably adjusted method of interpolation. Analogous to the fitting approach proposed by CFP we intend to approximate a rational function, defined over a higher-dimensional space, with the purpose of locating that point which fulfills the standard optimality condition. Practical difficulties relate, firstly, to the explicit definition of the set of non-overlapping hyperspaces over which we intend to interpolate, detailed in this section, and, secondly, in the application of a computationally efficient multidimensional interpolation algorithm that is capable of locating the desired optimal point within acceptable bounds of inaccuracy, the discussion of which is deferred to §4.6.6.

Paralleling the CFP approach, for the single threshold case, requires that we begin by pedantically defining non-overlapping threshold hypercubes applicable when extending the approach to a multiple threshold setting. An important feature of the approach taken is that each non-overlapping threshold hypercube, defined and constructed in the manner presented below, is conform the requirements stipulated in §4.4. Heuristically, each threshold candidate vector, $\boldsymbol{\theta}_g \in \boldsymbol{\Theta}$, will function as a support point upon which a single non-overlapping threshold hypercube, denoted $\mathbf{H}_g$, will be based, after which, its dimensionality will be

[29] We defer the discussion of this claim to § 4.6.3, while, the definition of $\mathbf{H}_g$ is defined later in this section.

assessed. This relates to a matter of practical difficulty since each hypercube[30] defines a subspace in $\Re^{k-1}$, yet, Θ is not a simple accumulation of $(k-1)-$dimensional hypercubes, but an accumulation of hypercubes with a maximum dimensionality of $(k-1)$ in $\Re^{k-1}$. Thus, if $k=4$, then the union of relevant hypercubes would constitute an accumulation of cubes, planes, lines and points. The focal point of the discussion relates to the precise definition of a non-overlapping threshold hypercube which, in turn, relies on a construct which we define as the set of inclusive boundaries. Let the row vector $\overline{\boldsymbol{\theta}}_g$ of $\mathbf{I}^{\boldsymbol{\Theta}}$ denote the mapped candidate threshold vector $\boldsymbol{\theta}_g$ of $\boldsymbol{\Theta}$:

Definition 1 (Set of Inclusive Boundaries):
Let each $\overline{\boldsymbol{\theta}}_g \in \mathbf{I}^{\boldsymbol{\Theta}}$ *define a support point and the set* $\left\{\overline{\boldsymbol{\theta}}_{\bar{g}} : \bar{g} \neq g, \overline{\boldsymbol{\theta}}_{\bar{g}} \in \mathbf{I}^{\boldsymbol{\Theta}}\right\}$ *denote all contiguous points in* $\mathbf{I}^{\boldsymbol{\Theta}}$ *that lie in a direction parallel to, and away from the origin of, a Cartesian coordinate system in* $\Re^{k-1}$. *Then the set of inclusive boundaries is defined as the collection of lines connecting* $\boldsymbol{\theta}_g \in \boldsymbol{\Theta}$ *with the set of contiguous gridpoints* $\left\{\boldsymbol{\theta}_{\bar{g}} : \bar{g} \neq g, \boldsymbol{\theta}_{\bar{g}} \in \boldsymbol{\Theta}\right\}$, *whereby all* $\boldsymbol{\theta}_{\bar{g}} \in \boldsymbol{\Theta}$, *for* $\bar{g} \neq g$, *are excluded from the set.*

Given this definition, $\mathbf{H}_g$ may be defined as:

Definition 2 (Non-overlapping Threshold Hypercube):
Consider the gridpoint $\boldsymbol{\theta}_g$ *and suppose that* n *boundaries are defined in the set of inclusive boundaries with which it is associated. Then* $\mathbf{H}_g$ *denotes the* n*-dimensional non-overlapping threshold hypercube, a subspace of* $\Re^{k-1}$, *defined as the union of all interior points of the* n*-dimensional hypercube that includes* $\left\{\boldsymbol{\theta}_{\bar{g}} : \bar{g} \neq g, \boldsymbol{\theta}_{\bar{g}} \in \boldsymbol{\Theta}\right\}$ *and the set of inclusive boundaries.*

Logically, the number of inclusive boundaries that comprise the set, defined above, for a given support point is equal to the dimensionality of the non-overlapping threshold hypercube they implicitly define. Practically, however, the dimensionality of the hypercube is deduced by contrasting the matrix $\overline{\mathbf{H}}_g = \left(\overline{\boldsymbol{\theta}}_g \otimes \boldsymbol{\iota}_{k-1}\right) + \mathbf{I}_{k-1}$ with $\mathbf{I}^{\boldsymbol{\Theta}}$. Since the row vectors of $\overline{\mathbf{H}}_g$ need to correspond with equivalent row vectors in $\mathbf{I}^{\boldsymbol{\Theta}}$ to comply with the requirements of §4.4, the

[30] For elucidatory purposes we use the term hypercube loosely here.

dimensionality of $\mathbf{H}_g$, is deducible from the number of rows in the set $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$. Hence, if $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$ is empty then $\mathbf{H}_g$ is a point in $\Re^{k-1}$, if $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$ constitutes a single element $\mathbf{H}_g$ is a line, and so on, until, finally, if the intersection has $k-1$ row vectors then $\mathbf{H}_g$ is a $(k-1)$-dimensional hypercube in $\Re^{k-1}$. Suppose that $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right)$ denotes the number of row vectors in this intersection then $2^{rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right)}$ vertices define the hypercube of interest, while, strictly, only $\boldsymbol{\theta}_g$ is included in the non-overlapping threshold hypercube.

The algorithm, delineated below, culminates in the construction of a matrix of candidate threshold (rows) vectors that will ultimately form the domain of the multidimensional interpolation-based optimization algorithm, discussed in the next section:

1. Construct $\boldsymbol{\Theta}$, the constrained grid, as described in §4.4.
2. For each $g \in \{1,2,\ldots,G\}$,
   a. Construct the matrix $\overline{\mathbf{H}}_g = \left(\overline{\boldsymbol{\theta}}_g \otimes \mathbf{1}_{k-1}\right) + \mathbf{I}_{k-1}$ and substitute all row vectors that are not elements of the intersection $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$, with $\overline{\boldsymbol{\theta}}_g$, and denote this (updated) matrix $\overline{\overline{\mathbf{H}}}_g$. Consequently, each row vector of $\overline{\overline{\mathbf{H}}}_g$ is in compliance with the requirements specified in §4.4.
   b. Map the (augmented) matrix given by $\left[\overline{\boldsymbol{\theta}}_g \mid \overline{\overline{\mathbf{H}}}_g\right]$, via the index number construct, back to the continuous threshold space.
   c. If $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$ is empty $\mathbf{H}_g = \boldsymbol{\theta}_g$.
   d. If $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$ is non-empty, compute a set of $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right)$ $\left(\max(q) \times (k-1)\right)$ matrices with row vectors that are the coordinates of a set of (equally spaced) discrete points, denoted $\boldsymbol{\theta}_{g,q}$ for $q \in \{1,2,\ldots,\max(q)\}$, lying on the line connecting $\boldsymbol{\theta}_g$ with each of the row vectors defined by the mapped rows of $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$ (Note that only the only vertex that may be included in each series of points is $\boldsymbol{\theta}_g$), where, by construction, if the *ith* element of the vector that results when subtracting $\overline{\boldsymbol{\theta}}_g$ from a particular row vector in $\overline{\overline{\mathbf{H}}}_g \cap \mathbf{I}^{\Theta}$ equals unity the line connecting these two gridpoints will be parallel to the *ith* coordinate axis.

e. Let $\bar{q} = \max(q)$ and suppose that the *ith* column of a particular matrix of coordinate vectors, described in d. above, has non-constant elements, then it is denoted by $\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$, while, if the *ith* row of $\overline{\overline{\mathbf{H}}}_g = \bar{\boldsymbol{\theta}}_g$, then $\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)} = \mathbf{1}_{\bar{q}}\theta_g^{(i)}$. The grid, containing the set of candidate threshold (row) vectors that will form the domain of the multidimensional interpolation-based optimization routine, is then given by the (horizontally concatenated) matrix $\boldsymbol{\Theta}_g = \left[\boldsymbol{\Theta}_{\mathbf{H}_g}^{(1)} \mid \boldsymbol{\Theta}_{\mathbf{H}_g}^{(2)} \mid \ldots \mid \boldsymbol{\Theta}_{\mathbf{H}_g}^{(k-1)}\right]$ where,

- If $k = 2$, $\quad \boldsymbol{\Theta}_{\mathbf{H}_g}^{(1)} = \boldsymbol{\theta}_{\mathbf{H}_g}^{(1)}$;
- If $k = 3$, $\quad \begin{aligned} \boldsymbol{\Theta}_{\mathbf{H}_g}^{(1)} &= \mathbf{1}_{\bar{q}} \otimes \boldsymbol{\theta}_{\mathbf{H}_g}^{(1)} \\ \boldsymbol{\Theta}_{\mathbf{H}_g}^{(2)} &= \boldsymbol{\theta}_{\mathbf{H}_g}^{(2)} \otimes \mathbf{1}_{\bar{q}} \end{aligned}$;
- For $k = 4$, $\quad \begin{aligned} \boldsymbol{\Theta}_{\mathbf{H}_g}^{(1)} &= \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \otimes \boldsymbol{\theta}_{\mathbf{H}_g}^{(1)} \\ \boldsymbol{\Theta}_{\mathbf{H}_g}^{(2)} &= \mathbf{1}_{\bar{q}} \otimes \boldsymbol{\theta}_{\mathbf{H}_g}^{(2)} \otimes \mathbf{1}_{\bar{q}} \\ \boldsymbol{\Theta}_{\mathbf{H}_g}^{(3)} &= \boldsymbol{\theta}_{\mathbf{H}_g}^{(3)} \otimes \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \end{aligned}$;
- For $k = 5$, $\quad \begin{aligned} \boldsymbol{\Theta}_{\mathbf{H}_g}^{(1)} &= \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \otimes \boldsymbol{\theta}_{\mathbf{H}_g}^{(1)} \\ \boldsymbol{\Theta}_{\mathbf{H}_g}^{(2)} &= \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \otimes \boldsymbol{\theta}_{\mathbf{H}_g}^{(2)} \otimes \mathbf{1}_{\bar{q}} \\ \boldsymbol{\Theta}_{\mathbf{H}_g}^{(3)} &= \mathbf{1}_{\bar{q}} \otimes \boldsymbol{\theta}_{\mathbf{H}_g}^{(3)} \otimes \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \\ \boldsymbol{\Theta}_{\mathbf{H}_g}^{(4)} &= \boldsymbol{\theta}_{\mathbf{H}_g}^{(3)} \otimes \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \otimes \mathbf{1}_{\bar{q}} \end{aligned}$

etc., where the pattern that defines the recurrence relation, above, is easily generalized to cases for which $k > 5$.

3) Finally, the grid that we desire to construct results from removing all non-unique row vectors from $\boldsymbol{\Theta}_g$

Once constructed, $\mathbf{H}_g$ will be sequentially activated and subjected to the application of a multidimensional interpolation-based optimization algorithm. Note that the grid, described above, is not Cartesian since each inclusive boundary will likely differ in length. Consequently, the set of grid points over which we compute the value of $\mathrm{RSS}_{\mathrm{g}}^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right)$ will be different from that upon which a simple GS, with fixed stepsize, is based. The next

section details a technical discussion, the underlying structure of which relates directly to the algorithmic form of the fitting approach that we espouse, presented in §4.7, on how Givens transformations are efficiently applied to QR factorizations in a multiple-threshold SETAR setting. We initiate the discussion by presenting a convenient reformulation of a general SETAR $(k;p;d)$ model that is amenable to the discussion at hand.

4.6. Generalizing the Fitting Approach to Allow for Higher-Dimensional Threshold Vectors

4.6.1. QR Factorizations in the Multiple-Threshold SETAR Context

The, previously defined, k-regime TAR model, (4.1), is re-expressed below such that the autoregressive processes in regimes that are an explicit function of $\boldsymbol{\theta}$ are separated from regimes that are independent thereof:

$$\Delta z_t = \begin{cases} \phi_0^{(j_1)} + \sum_{i=1}^{p^{(j_1)}} \phi_i^{(j_1)} f(z_{t-i}, \boldsymbol{\theta}) + \varepsilon_t^{(j_1)} & \text{if} \quad \theta^{(j_1-1)} \le v_{t-d} < \theta^{(j_1)} \\ \phi_0^{(j_2)} + \sum_{i=1}^{p^{(j_2)}} \phi_i^{(j_2)} z_{t-i} + \varepsilon_t^{(j_2)} & \text{if} \quad \theta^{(j_2-1)} \le v_{t-d} < \theta^{(j_2)} \end{cases} \tag{4.5}$$

where the sets of regime indices, denoted by j_1 and j_2, are mutually exclusive implying that if k_1 regimes are a function of $\boldsymbol{\theta}$ and k_2 are not then $k_1+k_2 = k$. Thus, if $k=3$; $v_{t-d} \equiv z_{t-d}$; $\phi_0^{(j_1)} = 0$; $j_1 = \{1;3\}$ and $j_2 = \{2\}$ where $-\theta^{(1)} = \theta^{(2)} = \theta$ and $-\theta^{(0)} = \theta^{(3)} = \infty$; and $f(z_{t-i}, \boldsymbol{\theta}) = z_{t-i} - \theta$ if $\theta \le z_{t-d}$, while, $f(z_{t-i}, \boldsymbol{\theta}) = z_{t-i} + \theta$ if $z_{t-d} < \theta$, then (4.5) has effectively been restricted to the symmetric Band-SETAR model on which the CFP procedure is based. Note that we do not desire to impose any specific structure on the AR processes of the model, instead, the reformulation, given in (4.5), is an expositional simplification designed to support the discussion on the proposed fitting approach. Next we discuss the QR factorization approach to solving LS problems in the multidimensional threshold vector framework. The approach is malleable to a number of relevant adaptations that we explicate after initializing the discussion on the methodology when it is free of *a priori* constraints. It is convenient to

restrict the continuous threshold space, Θ, to the set of candidate threshold locations, $\boldsymbol{\theta}_g$, defined as row vectors in $\boldsymbol{\Theta}$, for the moment.

Suppose $d \in \{1,\ldots,D\}$, where D is a plausible maximum value for the threshold delay, and L denotes a plausible a *priori* constraint on the maximum allowable lag order of each AR process, then, the $k \times L^k$ matrix $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{L^k}]$ where $\mathbf{p}_i = \left[p^{(k)}, p^{(k-1)}, \ldots, p^{(1)}\right]'$ defines all plausible lag order combinations for the AR processes in all regimes. Given compliance with the identifying restriction $k > 1$, then, conditional on a particular combination of AR lag orders, $\mathbf{p}_i$, the least squares approximation involves finding the k coefficient $p^{(j)}$-vectors, $\boldsymbol{\Phi}_g^{(j)}$, such that

$$\sum_{j=1}^{k} \left\| \mathbf{F0}_g^{(j)} \boldsymbol{\Phi}_g^{(j)} - \Delta \mathbf{z}_g^{(j)} \right\|_{2,N_g^{*(j)}} = \min_{\substack{1 \le g \le G, \\ 1 \le d \le D}} \left[\sum_{j=1}^{k} \inf_{\mathbf{v}_g^{(j)} \in \Re^{p^{(j)}}} \left\| \mathbf{F0}_g^{(j)} \mathbf{v}_g^{(j)} - \Delta \mathbf{z}_g^{(j)} \right\|_{2,N_g^{*(j)}} \right] \tag{4.6}$$

where $\|\|_{2,N_g^{*(j)}}$ denotes the Euclidean norm in $\Re^{N_g^{*(j)}}$. Assume $\mathbf{F0}_g^{(j)}$ is a full-column rank matrix for all j, $\mathbf{c} \in \Re^{p^{(j)}}$ and $\alpha \in \Re$ and consider the following equality

$$\left\| \mathbf{F0}_g^{(j)} \left(\boldsymbol{\Phi}_g^{(j)} + \alpha \mathbf{c} \right) - \Delta \mathbf{z}_g^{(j)} \right\|_2^2 = \left\| \mathbf{F0}_g^{(j)} - \Delta \mathbf{z}_g^{(j)} \right\|_2^2 + 2\alpha \mathbf{c}' \mathbf{F0}_g^{(j)'} \left(\mathbf{F0}_g^{(j)} \boldsymbol{\Phi}_g^{(j)} - \Delta \mathbf{z}_g^{(j)} \right) + \alpha^2 \left\| \mathbf{F0}_g^{(j)} \mathbf{c} \right\|_2^2$$

then, if $\boldsymbol{\Phi}_g^{(j)}$ solves this least squares problem, for the support point $\boldsymbol{\theta}_g$, we must have $\mathbf{F0}_g^{(j)'} \left(\mathbf{F0}_g^{(j)} \boldsymbol{\Phi}_g^{(j)} - \Delta \mathbf{z}_g^{(j)} \right) = 0$ for all j. Thus, given that $\mathbf{F0}_g^{(j)}$ is a full-column rank matrix, there is a unique LS solution $\hat{\boldsymbol{\Phi}}_g^{(j)}$ that solves the symmetric positive linear system $\mathbf{F0}_g^{(j)'} \mathbf{F0}_g^{(j)} \hat{\boldsymbol{\Phi}}_g^{(j)} = \mathbf{F0}_g^{(j)'} \Delta \mathbf{z}_g^{(j)}$ in each regime. This solution may be conveniently computed via a QR factorization[31]. Suppose that, for all j, an orthogonal matrix $\mathbf{Q}_g^{(j)} \in \Re^{N_g^{*(j)} \times N_g^{*(j)}}$ has been

[31] If $\mathbf{F}_g^{(j)}$ is not of full rank, or if its rank is in doubt, a QR factorization may be performed with column pivoting or a singular value decomposition (SVD).

computed such that $\mathbf{Q}_g^{(j)'}\mathbf{F0}_g^{(j)} = \mathbf{R}_g^{(j)} = \begin{bmatrix} \mathbf{R}_{g,1}^{(j)} \\ \mathbf{0} \end{bmatrix} \begin{matrix} p^{(j)} \\ N^{*(j)} - p^{(j)} \end{matrix}$ is upper triangular. Suppose $\mathbf{Q}_g^{(j)'}\Delta\mathbf{z}_g^{(j)} = \begin{bmatrix} \mathbf{d}_{g,1}^{(j)} \\ \mathbf{d}_{g,2}^{(j)} \end{bmatrix} \begin{matrix} p^{(j)} \\ N_g^{*(j)} - p^{(j)} \end{matrix}$, then

$$\left\|\mathbf{F0}_g^{(j)}\boldsymbol{\Phi}_g^{(j)} - \Delta\mathbf{z}_g^{(j)}\right\|_2^2 = \left\|\mathbf{Q}_g^{(j)'}\mathbf{F0}_g^{(j)}\boldsymbol{\Phi}_g^{(j)} - \mathbf{Q}_g^{(j)'}\Delta\mathbf{z}_g^{(j)}\right\|_2^2 = \left\|\begin{bmatrix} \mathbf{R}_{g,1}^{(j)} \\ \mathbf{0} \end{bmatrix}\boldsymbol{\Phi}^{(j)} - \begin{bmatrix} \mathbf{d}_{g,1}^{(j)} \\ \mathbf{d}_{g,2}^{(j)} \end{bmatrix}\right\|$$
$$= \left\|\mathbf{R}_{g,1}^{(j)}\boldsymbol{\Phi}_g^{(j)} - \mathbf{d}_{g,1}^{(j)}\right\|_2^2 + \left\|\mathbf{d}_{g,2}^{(j)}\right\|_2^2$$

and, clearly, if $Rank\left(\mathbf{F0}_g^{(j)}\right) = Rank\left(\mathbf{R}_{g,1}^{(j)}\right) = p^{(j)}$ then $\hat{\boldsymbol{\Phi}}_g^{(j)}$ is defined by the upper-triangular system $\mathbf{R}_{g,1}^{(j)}\hat{\boldsymbol{\Phi}}_g^{(j)} = \mathbf{d}_{g,1}^{(j)}$, letting $\hat{\boldsymbol{\varepsilon}}_g^{(j)}\left(\boldsymbol{\theta}_g, d, p^{(j)}\right) = \Delta\mathbf{z}_g^{(j)} - \mathbf{F0}_g^{(j)}\hat{\boldsymbol{\Phi}}_g^{(j)}$, where $p^{(j)} \in \mathbf{p}_i$, the residual sum of squares for the model as a whole, associated with $\boldsymbol{\theta}_g$, $\mathbf{p}_i$ and d, is defined as

$$RSS_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right) = \sum_{j=1}^{k}\hat{\boldsymbol{\varepsilon}}_g^{(j)}\left(\boldsymbol{\theta}_g, d, p^{(j)}\right)'\hat{\boldsymbol{\varepsilon}}_g^{(j)}\left(\boldsymbol{\theta}_g, d, p^{(j)}\right) = \sum_{j=1}^{k}\left\|\mathbf{d}_{g,2}^{(j)}\right\|_2^2$$

which is, subsequently, utilized to determine the best-fitting lag order for the given active candidate threshold vector. Let the Akaike information criteria (AIC)[32], for the TAR model as a whole, given $\boldsymbol{\theta}_g$, $\mathbf{p}_i$, and d, be formulated as

$$AIC_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right) = \left[N^* \ln\left(RSS_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right)/N^*\right)\right] + 2\boldsymbol{\iota}_k'\mathbf{p}_i$$

For a given active candidate threshold vector, $\boldsymbol{\theta}_g$, let $\tilde{\mathbf{p}}_{\mathbf{H}_g} = \underset{1\le i\le L^k}{\arg\min}\, AIC_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right)$ and $AIC_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d\right) = AIC_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \tilde{\mathbf{p}}_{\mathbf{H}_g}\right)$, then $\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{\mathbf{p}}\right) = \underset{\substack{1\le g\le G,\\ 1\le d\le D}}{\arg\min}\, AIC_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d\right)$. The fitting approach, stipulated in §4.7, requires that the orthogonal matrix $\mathbf{Q}_1^{(j)}$ and the (unique) Cholesky factor of $\mathbf{F0}_1^{(j)'}\mathbf{F0}_1^{(j)}$, $\mathbf{R}_1^{(j)}$, be computed, after which, Givens transformations will be utilized to update these matrices for $g > 1$.

[32] This measure is suggested by Tong (1983).

4.6.2. Updating QR Factorizations: $\boldsymbol{\theta}_g \rightarrow \boldsymbol{\theta}_{g+1}$

The grid, $\boldsymbol{\Theta}$ (see §4.4), is so defined as to facilitate the effective implementation of QR factorization updating algorithms. Consequently, considering the next candidate threshold vector, $\boldsymbol{\theta}_{g+1}$, in $\boldsymbol{\Theta}$ implies that only two contiguous regimes will be affected by the redistribution of cases. A number of viable approaches exist when tackling the problem of updating the factorizations computed conditional on $\boldsymbol{\theta}_g$. The approach that we have opted for may be simply stated as[33]:

a) If $N_{g+1}^{*(j)} < N_g^{*(j)}$, consider the QR factorization $\mathbf{Q}_g^{(j)}\mathbf{R}_g^{(j)} = \mathbf{F0}_g^{(j)}$, given the candidate threshold $\boldsymbol{\theta}_g$, when the first row of $\mathbf{F0}_g^{(j)}$ has been deleted[34]. Specifically, the computation of the QR factorization of the submatrix $\overline{\mathbf{F0}}_g^{(j)}$ is desired in $\mathbf{F0}_g^{(j)} = \begin{bmatrix} \mathbf{w}' \\ \overline{\mathbf{F0}}_g^{(j)} \end{bmatrix} \begin{matrix} 1 \\ N_g^{*(j)} - 1 \end{matrix}$. Let $\mathbf{q}_g^{(j)'}$ denote the first row of $\mathbf{Q}_g^{(j)}$ and compute the Givens rotations $\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_{N_g^{*(j)}-1}$ such that $\mathbf{G}_1' \cdots \mathbf{G}_{N_g^{*(j)}-1}' \mathbf{q}_g^{(j)} = \alpha e_1$ where e_1 is the first column of $\mathbf{I}_{N_g^{*(j)}}$ and $\alpha = \pm 1$. Note that $\mathbf{G}_1' \cdots \mathbf{G}_{N_g^{*(j)}-1}' \mathbf{R}_g^{(j)} = \begin{bmatrix} \mathbf{v}' \\ \overline{\mathbf{R}}_g^{(j)} \end{bmatrix} \begin{matrix} 1 \\ N_g^{*(j)} - 1 \end{matrix}$ is upper Hessenberg and that $\mathbf{Q}_g^{(j)} \mathbf{G}_{N_g^{*(j)}-1} \cdots \mathbf{G}_1 = \begin{bmatrix} \alpha & 0 \\ 0 & \overline{\mathbf{Q}}_g^{(j)} \end{bmatrix}$, where $\overline{\mathbf{Q}}_g^{(j)} \in \Re^{\left(N_g^{*(j)}-1\right) \times \left(N_g^{*(j)}-1\right)}$ is orthogonal. Then since $\mathbf{F0}_g^{(j)} = \left(\mathbf{Q}_g^{(j)} \mathbf{G}_{N_g^{*(j)}-1} \cdots \mathbf{G}_1\right)\left(\mathbf{G}_1' \cdots \mathbf{G}_{N_g^{*(j)}-1}' \mathbf{R}_g^{(j)}\right) = \begin{bmatrix} \alpha & 0 \\ 0 & \overline{\mathbf{Q}}_g^{(j)} \end{bmatrix} \begin{bmatrix} \mathbf{v}' \\ \overline{\mathbf{R}}_g^{(j)} \end{bmatrix}$ it is easy to see that the desired QR factorization is given by $\overline{\mathbf{F0}}_g^{(j)} = \overline{\mathbf{Q}}_g^{(j)} \overline{\mathbf{R}}_g^{(j)}$ which is the QR factorization of $\mathbf{F0}_{g+1}^{(j)}$.

---

[33] As argued by Eğecioğlu and Srinivasan (1995), if a large number of cases are transferred from one regime to another by moving from $\boldsymbol{\theta}_g$ to $\boldsymbol{\theta}_{g+1}$, as is often the case when a TAR model is estimated on a large dataset, or when the set of candidate thresholds is restricted to being a subset of the order statistics of the delay variable, it maybe preferable to use a Household reflector as opposed to Givens rotations. Both solutions are preferred over the solution of the normal equations for linear least squares problems (Wilkinson (1965), Golub and van Loan (1990), and Golub and Ortega (1993)).

[34] The procedure is similar when an arbitrary row of $F_0^{(j)}$ is deleted.

b) If $N_{g+1}^{*(j)} > N_g^{*(j)}$, consider the QR factorization $\mathbf{Q}_g^{(j)}\mathbf{R}_g^{(j)} = \mathbf{F0}_g^{(j)} \in \Re^{N_g^{*(j)} \times p^{(j)}}$, when $\boldsymbol{\theta}_g$ is the support point of the active hypercube, and suppose that the regressor matrix $\mathbf{F0}_g^{(j)}$ is appended with another row i.e. $\overline{\mathbf{F0}}_g^{(j)} = \begin{bmatrix} \mathbf{w}' \\ \mathbf{F0}_g^{(j)} \end{bmatrix}$ where $\mathbf{w} \in \Re^{p^{(j)}}$. Note that $diag\left(1, \mathbf{Q}_g^{(j)'}\right)\overline{\mathbf{F0}}_g^{(j)} = \begin{bmatrix} \mathbf{w}' \\ \mathbf{R}_g^{(j)} \end{bmatrix}$ is upper Hessenberg. Consequently, Givens rotations $\mathbf{J}_1, \ldots, \mathbf{J}_{p^{(j)}}$ may be determined resulting in the upper triangular matrix $\overline{\mathbf{R}}_g^{(j)} = \mathbf{J}'_{p^{(j)}} \cdots \mathbf{J}'_1 \begin{bmatrix} \mathbf{w}' \\ \mathbf{R}_g^{(j)} \end{bmatrix}$. Since $\overline{\mathbf{Q}}_g^{(j)} = diag\left(1, \mathbf{Q}_g^{(j)}\right)\mathbf{J}_1 \cdots \mathbf{J}_{p^{(j)}}$ it follows that the desired QR factorization is given by $\overline{\mathbf{F0}}_g^{(j)} = \overline{\mathbf{Q}}_g^{(j)}\overline{\mathbf{R}}_g^{(j)}$.

The next section circumstantiates how the CFP result, that $RSS_g^{(j)}\left(\mathbf{H}_g, d, p^{(j)}\right)$ is both continuous and rational, may be malleated into a form that is amenable to the multiple threshold TAR context, in which we are interested. Consequently, Larkin's (1967) rational interpolation methodology is discussed once we have delineated how rank-one Givens corrections may be utilized to update QR factorizations when extending the search set from $\boldsymbol{\theta}_g$ to $\mathbf{H}_g$. Finally, we propose a multidimensional interpolation-based optimization routine designed to expediently pinpoint the location in $\mathbf{H}_g$ that minimizes $RSS_g^{(1,2,\ldots,k)}\left(\mathbf{H}_g, d, \mathbf{p}_i\right)$.

4.6.3. The Standard Optimality Condition in $\mathbf{H}_g$

If the independent variables are not an explicit function of $\boldsymbol{\theta}$ the least squares requirement, stipulated previously, or otherwise stated, maximization of the likelihood function, has been achieved. Alternatively, if $\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}$ is non-empty (see §4.5) and the order statistics of the delay variable do not suffice as candidate threshold vectors, since the independent variables depend explicitly on $\boldsymbol{\theta}$, then, dependent on the functional form of the sub-models in the relevant regimes, appropriate updates of the QR factorizations will be required.

Consider the following decompositions of plausible functional forms of the model in regime j: $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}$ and $\mathbf{F}_g^{(j)} = \mathbf{U}_g^{(j)} \circ \mathbf{F0}_g^{(j)}$ where $\mathbf{U}_g^{(j)}$ is a correction matrix the specific

form of which is dependent on the functional form present in regime j and $\circ$ denotes the Hadamard product[35]. For the additive functional form[36],

$$\mathbf{U}_g^{(j)} = \left[c_1^{(j)}, c_2^{(j)}, \ldots, c_{N_g^{*(j)}}^{(j)}\right]\left[\theta_{\mathbf{H}_g}^{(j)}, \theta_{\mathbf{H}_g}^{(j)}, \ldots, \theta_{\mathbf{H}_g}^{(j)}\right] \tag{4.8}$$

where $\left[c_1^{(j)}, c_2^{(j)}, \ldots, c_{N_g^{*(j)}}^{(j)}\right]$ is a vector of scalars[37] and $\theta_{\mathbf{H}_g}^{(j)} \in \boldsymbol{\theta}_{\mathbf{H}_g}^{(j)}$, for very large $\bar{q}$ (see §4.5). Clearly, a plethora of possibilities underlie this formulation of the correction matrix.

An interesting example of a correction matrix for the multiplicative functional form of a three-regime SETAR model could be given by

$$\begin{aligned}\mathbf{U}_g^{(j)} = &\left[\boldsymbol{\delta}^{(1)} \otimes \boldsymbol{\iota}'_{p^{(1)}}\right] I\left(z_{t-d} < \theta^{(1)}\right) + \left[\boldsymbol{\iota}_{N_g^{*(2)}} \otimes \boldsymbol{\iota}'_{p^{(2)}}\right] I\left(\theta^{(1)} \le z_{t-d} \le \theta^{(2)}\right) + \\ &\left[\boldsymbol{\delta}^{(3)} \otimes \boldsymbol{\iota}'_{p^{(3)}}\right] I\left(z_{t-d} > \theta^{(2)}\right)\end{aligned} \tag{4.9}$$

where $\otimes$ denotes the Kronecker product, $\boldsymbol{\delta}^{(j)}$ is the $N_g^{*(j)}$-vector $\left[\delta_{(1)}^{(j)}, \delta_{(2)}^{(j)}, \ldots, \delta_{(N_g^{*(j)})}^{(j)}\right]$ where $\delta_{(i)}^{(3)} = \left|z_{(i)} - \theta^{(2)}\right|$ and $\delta_{(i)}^{(1)} = \left|z_{(i)} - \theta^{(1)}\right|$ with $z_{(i)}$ being the *ith* order statistic of the delay variable[38]. This model has the capacity to accelerate/decelerate reversion of the process to its equilibrium state, within an outer-regime, depending on the size of the discrepancy between the value of the delay variable and the relevant threshold while maintaining an AR process in the inner-regime. It is evident that, conceptually, the decomposition of the regressor matrix in this manner facilitates an understanding of the underlying dynamics of models of this type. Models entailing multiplicative functional forms are, as yet, in an embryonic stage of development. Consequently, our attention is directed towards the simpler case of additive functional forms when discussing the implementation of Givens corrections, alluded to previously, of the QR decomposition when $\boldsymbol{\theta} = \boldsymbol{\theta}_g$. To this end, the following theorem, the proof of which may be found in the appendix of CFP, will form the foundation of the proposed optimization routine:

---

[35] The Hadamard product is defined as $A \circ B = (a_{ij} b_{ij})$ if $A = (a_{ij})$ and $B = (b_{ij})$ are $m \times n$ matrices of the same order. Thus, the Hadamard product is an $m \times n$ matrix with the *ij-th* element equal to $a_{ij} b_{ij}$.

[36] Replicated from §4.2 for purposes of clarity.

[37] Note the implications of this general formulation.

[38] To avoid division by zero the interval spanned by the inner regime includes the thresholds.

CFP Theorem (Rationality of $RSS_g^{(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$):

If the $N_g^{*(j)} \times p^{(j)}$ $\left(N_g^{*(j)} \geq p^{(j)}\right)$ *regressor matrix* $\mathbf{F}_g^{(j)}$ *in* $\inf_{\mathbf{v} \in \Re^{p^{(j)}}} \left\| \mathbf{F}_g^{(j)} \mathbf{v}^{(j)} - \Delta \mathbf{z}_g^{(j)} \right\|_{2, N^{*(j)}}$ *is a first degree polynomial matrix* $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}$, *where* $\mathbf{U}_g^{(j)} = \mathbf{C}\theta_g^{(i)}$ *with* $\theta_g^{(i)} \in \boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$, *for very large* $\bar{q}$, *and* $\mathbf{C}$, *a matrix of rank one whose ith row is* $(c_i, \ldots, c_i)$ *with* c_i *constant, then the sum of squared residuals* $\left\| \hat{\mathbf{e}}_g^{(j)} \right\|_2^2$ *is a rational function of degree type* $(4,2)$ *provided* $\mathbf{F}_g^{(j)}$ *is a full-column rank matrix.*

Thus, $\left\| \hat{\mathbf{e}}_g^{(j)} \right\|_2^2$ is a rational function defined over $\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$, for large $\bar{q}$, and is of degree type $(4,2)$ i.e. $RSS_g^{(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right) \equiv \Psi_g^{4,2(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$. Consider the non-overlapping threshold hypercube $\mathbf{H}_g$ and a combination of regimes, for which it is true that for some $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)}$ and for others $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}$, then, since $\left(\mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}\right)' \left(\mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}\right)$ is nonsingular (i.e. the denominator of $\Psi_g^{4,2(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$ never vanishes) and the RSS function is invariant over $\mathbf{H}_g$ when $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)}$, the *RSS* function for the model as a whole is continuous and given by

$$RSS_g^{(1,2,\ldots,k)}\left(\mathbf{H}_g, d, \mathbf{p}_i\right) = \sum_{j_1=1}^{k_1} \Psi_g^{4,2(j_1)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right) + \sum_{j_2=1}^{k_2} RSS_g^{(j_2)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$$

which results in the optimal location $\boldsymbol{\theta}_g^* = \arg\min_{\boldsymbol{\theta} \in \mathbf{H}_g} RSS_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}, d, \mathbf{p}_i\right)$. Thus, $\boldsymbol{\theta}_g^*$ denotes the location of the candidate threshold vector which minimizes the *RSS* function of the entire model when searching over $\mathbf{H}_g$. The search routine designed to locate this point has been detailed in §4.6.6, while Larkin's (1967) rational interpolation method, upon which this routine is based, is presented subsequent to the discussion on rank-one Givens corrections.

4.6.4. Updating QR Factorizations: $\theta_g^{(j)} \rightarrow \theta_{\mathbf{H}_g,q}^{(i)}$

Suppose that the QR factorization $\mathbf{Q}_g^{(j)}\mathbf{R}_g^{(j)} = \mathbf{F0}_g^{(j)} \in \Re^{N_g^{*(j)} \times p^{(j)}}$ has been computed and that the QR factorization of the first degree polynomial matrix $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)} = \overline{\mathbf{Q}}_g^{(j)}\overline{\mathbf{R}}_g^{(j)}$, where $\mathbf{U}_g^{(j)}$ is a function of $\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$, is required. Observe that $\mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)} = \mathbf{Q}_g^{(j)}\left(\mathbf{R}_g^{(j)} + \mathbf{w}\theta_{\mathbf{H}_g,q}^{(i)}\right)$ where $\mathbf{w} = \mathbf{Q}_g^{(j)'}\mathbf{C}$ and $\theta_{\mathbf{H}_g,q}^{(i)} \in \boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$. Suppose that a set of Givens rotations $\mathbf{J}_{p^{(j)}-1},\ldots,\mathbf{J}_2,\mathbf{J}_1$ have been computed such that $\mathbf{J}_1'\cdots\mathbf{J}_{p^{(j)}-1}'\mathbf{w} = \pm\|\mathbf{w}\|_2\mathbf{e}_1$, $\mathbf{e}_1$ being the first column of $\mathbf{I}_{N_g^{*(j)}}$, then it can be shown that $\mathbf{h} = \mathbf{J}_1'\cdots\mathbf{J}_{p^{(j)}-1}'\mathbf{R}_g^{(j)}$ is upper Hessenberg. Consequently, $\left(\mathbf{J}_1'\cdots\mathbf{J}_{p^{(j)}-1}'\right)\left(\mathbf{R}^{(j)} + \mathbf{w}\theta_{\mathbf{H}_g,q}^{(i)}\right) = \mathbf{h} \pm \|\mathbf{w}\|_2\mathbf{e}_1\theta_{\mathbf{H}_g,q}^{(i)} = \overline{\mathbf{h}}$ is also upper Hessenberg. Givens rotations $\mathbf{G}_1,\ldots,\mathbf{G}_{p^{(j)}-1}$ may then be computed such that $\mathbf{G}_{p^{(j)}-1}'\cdots\mathbf{G}_1'\overline{\mathbf{h}} = \overline{\mathbf{R}}_g^{(j)}$, from which $RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d, \mathbf{p}_i\right)$ maybe easily be deduced, is upper triangular[39]. The $\overline{q}$ pairs of points given by $\left(\theta_{\mathbf{H}_g,q}^{(i)}, RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d, \mathbf{p}_i\right)\right)$ will function as interpolating points in the rational interpolation algorithm delineated in the next section.

4.6.5. Rational Interpolation

If a set of $\overline{q} = m + n + 1$ pairs of points $\left(\theta_{\mathbf{H}_g,q}^{(i)}, RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d, \mathbf{p}_i\right)\right)$ for $1 \le q \le \overline{q}$ are given, where, by construction, the $\overline{q}$ $\theta_{\mathbf{H}_g,q}^{(i)}$'s are points on a line, that is both a subset of $\mathbf{H}_g$ and lies in a direction parallel to the *ith* threshold axis, and the RSS function of regime *j* is a functional of $\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$, then the rational interpolation problem is defined as the task of determining a rational function $\Psi_g^{m,n(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right) = p_m\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right) / q_n\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$, where $p_m\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$ and $q_n\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$ are polynomials of degree m and n, respectively, such that

[39] Note that, from the perspective of the proposed algorithm, it is unnecessary to determine $\overline{\mathbf{Q}}_g^{(j)}$ when implementing rank-one corrections.

$\Psi_g^{(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right) = RSS_g^{(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$ for all $1 \le q \le \bar{q}$[40]. The CFP theorem, above, has shown that the RSS function is rational of degree type $(4,2)$ i.e. $m = 4$ and $n = 2$, subsequently each interpolation will require 7 pairs of points. Further, let $\bar{\theta}_{\mathbf{H}_g,h}^{(i)}$ for $h = 1,2,\ldots,\max(h)$ denote a sequence of discrete evaluation points where $\bar{\theta}_{\mathbf{H}_g,h}^{(i)} \in \boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}$, when $\bar{q}$ is very large.

By recasting Stoer's (1961) tabular methods, into a slightly simpler form, Larkin (1967) successfully integrated the classical Newton-Neville-Aitken techniques for polynomial interpolation and the Thiele continued fraction method (see Thiele (1909), Aitken (1932) and Neville (1934)) into an algorithm suited to the task of efficiently numerically interpolating a rational function[41]. Larkin's (1967) methodology entails the construction of a matrix of interpolant values $\left\{f_{q,k}\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)}\right)\right\}$ for $q = 1,2,\ldots,\bar{q}$ and $k = 1,2,\ldots,\max(k)$ each one possessing the property that $f_{q,k}\left(\theta_{\mathbf{H}_g,r}^{(i)}, d, \mathbf{p}_i\right) = \Psi_g^{4,2(j)}\left(\theta_{\mathbf{H}_g,r}^{(i)}, d, \mathbf{p}_i\right)$ for $q \le r \le q+k$[42]. Utilizing the "triangle" (or Neville-Aitken formula) and "rhombus" rules, $\Psi_g^{4,2(j)}\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)}, d, \mathbf{p}_i\right)$ for $h = 1,2,\ldots,\max(h)$ are conveniently computed in a $((m+n)\times(m+n))$ matrix, where the columnar progression implied by the rules results in a matrix of function values on and above the main skew diagonal. The "triangle rule" may be formulated as

$$f_{q,k}\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)}\right) = \frac{\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)} - \theta_{\mathbf{H}_g,q}^{(i)}\right) f_{q+1,k+1}\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)}\right) + \left(\theta_{\mathbf{H}_g,q+k}^{(i)} - \bar{\theta}_{\mathbf{H}_g,h}^{(i)}\right) f_{q,k-1}\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)}\right)}{\theta_{\mathbf{H}_g,q+k}^{(i)} - \theta_{\mathbf{H}_g,q}^{(i)}} \tag{4.10}$$

whereas the "rhombus rule" is given by

[40] Clearly, a rational interpolant may not exist when fixing the degree types of the numerator and denominator polynomials, implying the potential existence of a point set with a degenerate configuration.

[41] An eloquent algorithm, due to Jacobi, yields exact determinantal formulae for the coefficients of the denominator and numerator polynomials by directly solving the set of linear equations $p_m\left(\theta_{\mathbf{H}_g,q}^{(i)}, d, \mathbf{p}_i\right) - \Psi_g^{m,n(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d, \mathbf{p}_i\right) q_n\left(\theta_{\mathbf{H}_g,q}^{(i)}, d, \mathbf{p}_i\right) = 0$ for $1 \le q \le \bar{q}$. The instability of the technique, however, renders it unsuitable for the application at hand.

[42] Except under special circumstances.

$$f_{q,k}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)= f_{q+1,k-2}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)+ \frac{\theta^{(i)}_{\mathbf{H}_g,q+k}-\theta^{(i)}_{\mathbf{H}_g,q}}{\frac{\bar{\theta}^{(i)}_{\mathbf{H}_g,h}-\theta^{(i)}_{\mathbf{H}_g,q}}{f_{q+1,k-1}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)-f_{q+1,k-2}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)}+\frac{\theta^{(i)}_{\mathbf{H}_g,q+k}-\bar{\theta}^{(i)}_{\mathbf{H}_g,h}}{f_{q,k-1}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)-f_{q+1,k-2}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)}} \tag{4.11}$$

where $f_{q,0}=\Psi^{4,2}_{g^{(j)}}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,q},d,\mathbf{p}_i\right)$ and $f_{q,k}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h}\right)$ corresponds with the function value entry in row q of the *kth* column of the interpolant value matrix. Following the zigzag pattern[43] demanded by the approach (see Larkin (1967)) the first three columns are computed using (4.10), after which all successive columns are computed using (4.11). The function value on the main skew diagonal *and* the last column is the desired quantity, i.e. $\Psi^{4,2}_{g^{(j)}}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h},d,\mathbf{p}_i\right)$. Finally, the optimal interpolant value is given by $\tilde{\bar{\theta}}^{(i)}_{\mathbf{H}_g}=\underset{1\le h\le \max(h)}{\arg\min}\,\Psi^{4,2}_{g^{(j)}}\left(\bar{\theta}^{(i)}_{\mathbf{H}_g,h},d,\mathbf{p}_i\right)$[44].

4.6.6. Multidimensional Interpolation-based Optimization

This section pedagogically details a novel multidimensional interpolation-based optimization routine designed to take advantage of the well-behavedness of $\Psi^{4,2}_{g^{(j)}}\left(\bar{\boldsymbol{\theta}}^{(i)}_{\mathbf{H}_g},d,\mathbf{p}_i\right)$ in $\mathbf{H}_g$, a claim that is poignantly illustrated in figure 1, and discussed below, while simultaneously utilizing matrix constructs that allow for the expedient determination of the optimal location of the thresholds given a multiple-threshold SETAR framework. Simply stated, the routine involves breaking the multidimensional interpolation problem into a succession of one-dimensional interpolations; hence we begin by pedantically presenting the one-dimensional interpolation routine.

[43] The set of intermediate rational function values is computed following the pattern $(0,1)$, $(0,2)$ and $(0,3)$, by using equation (4.10), after which the function values for the pattern $(1,3)$, $(1,4)$ and $(2,4)$ are computed using equation (4.11).

[44] The polynomials may be efficiently evaluated using Horner's method (see Séroul (2000)).

a) Interpolating $\Psi_g^{4,2}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}^{(i)}, d, \mathbf{p}_i\right)$

The optimization routine, presented below, will specify a $(2\times(k-1))$ matrix of line end-point coordinates and *rows*$\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\mathbf{e}}\right)$ when it repetitively calls the following function:

1. Construct a $(\bar{q}\times(k-1))$ matrix of threshold locations, where each row vector specifies the location denoted by $\boldsymbol{\theta}_{\mathbf{H}_g,q}$, along the line connecting the end-point coordinates specified as an input when this function is called. Note the requirement that $\theta_{\mathbf{H}_g,q}^{(i)} \in \mathbf{H}_g$, for $q = 1,2,\ldots,\bar{q}$, and that $\theta_{\mathbf{H}_g,q}^{(i)}$ is the *ith* column of this matrix.
2. Determine the direction of the line i.e. *i*. This is easily done since the *ith* column of the matrix defined in step 1 will specify a series of elements that range from $\boldsymbol{\theta}_{\mathbf{H}_g,1}^{(i)}$ to $\boldsymbol{\theta}_{\mathbf{H}_g,\bar{q}}^{(i)}$ whereas all other columns will contain elements that are invariant through the $\bar{q}$ rows.
3. Construct a $(\max(h)\times(k-1))$ matrix containing the coordinates of (equally-spaced) evaluation points, where the *hth* element of the *ith* column of this matrix, denoted $\boldsymbol{\theta}_{\mathbf{H}_g,h}^{(i)}$, are also elements on the line formed by $\theta_{\mathbf{H}_g,q}^{(i)} \in \mathbf{H}_g$ when $\bar{q}$ is set to being very large.
4. Consider each row in the matrix constructed in step 1 and the direction *i* determined in step 2:
   a. Suppose that regime $\bar{j}$ is an explicit function of $\theta_{\mathbf{H}_g}^{(i)}$. Update $\mathbf{R}_g^{(j)}$, the Cholesky factor computed given the support point $\boldsymbol{\theta}_g$, using rank-one Givens updates (see §4.6.4).
   b. It is convenient to construct the $(\bar{q}\times L)$ matrix of RSS values by applying Givens rotations, for $p^{(j)} = 1,2,\ldots,L$, discussed in §4.6.7 for each threshold location $\boldsymbol{\theta}_{\mathbf{H}_g,q}^{(i)} \in \mathbf{H}_g$.

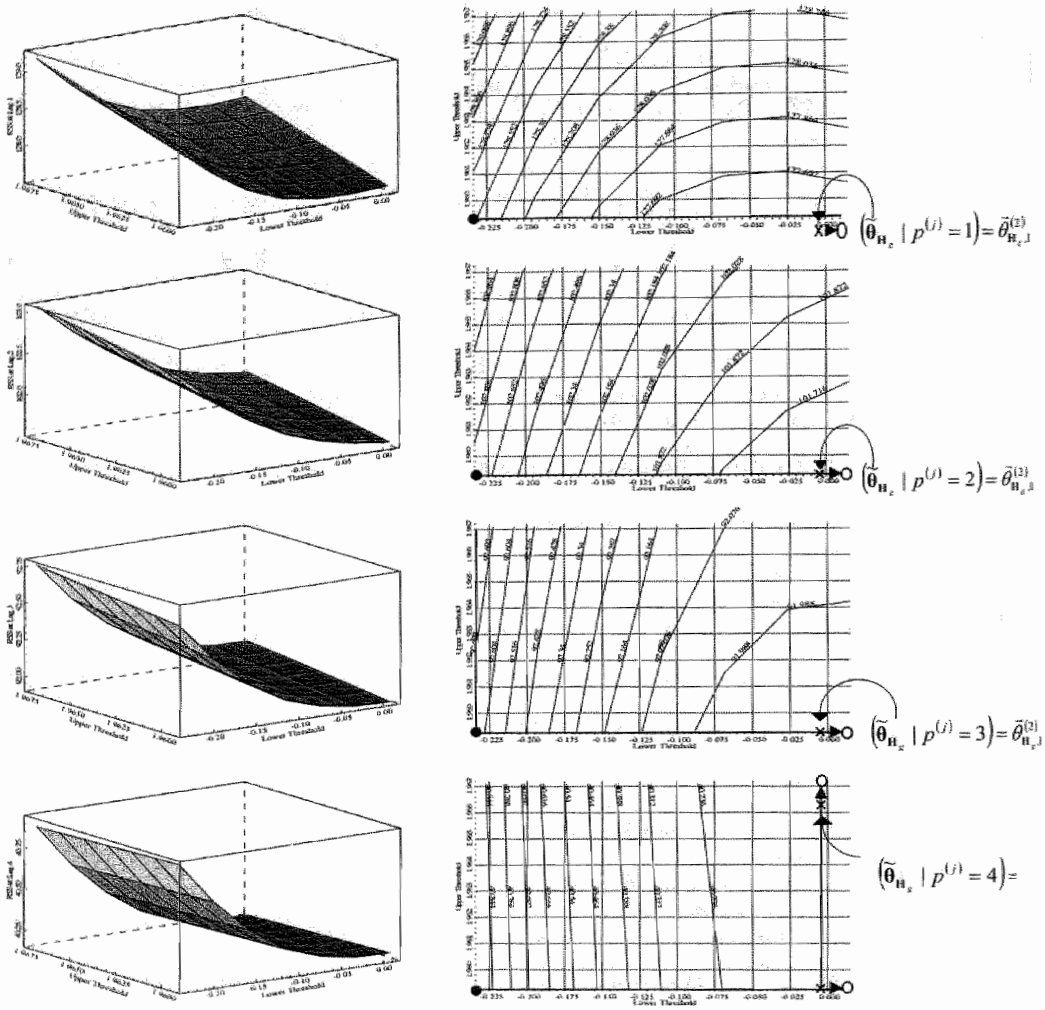

Figure 1. A set of 3-d scatter surfaces illustrating typical forms of the RSS function in various non-overlapping threshold hypercubes.
$RSS_{g}^{(1,2,3)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(j)}, d, \mathbf{p}_i\right)$ in $\mathbf{H}_g$, illustrated as a 3-d scatter surface and a 2-d contour of this 3-d scatter surface. The route taken by the proposed optimization routine is illustrated on the contour plot. Parameter settings are equivalent to those employed in the Monte Carlo simulations, see §4.8, of the asymmetric Band-SETAR process, where additionally: $N = 50$ and $\sigma_{\varepsilon}^2 = 0.9$.

$$\begin{bmatrix} RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d \mid q=1, p^{(j)}=1\right) & RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d \mid q=1, p^{(j)}=2\right) & \cdots & RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d \mid q=1, p^{(j)}=L\right) \\ RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d \mid q=2, p^{(j)}=1\right) & & \ddots & \vdots \\ \vdots & & & \\ RSS_g^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d \mid q=\bar{q}, p^{(j)}=1\right) & & \cdots & RSS^{(j)}\left(\theta_{\mathbf{H}_g,q}^{(i)}, d \mid q=\bar{q}, p^{(j)}=L\right) \end{bmatrix}$$

5. Utilizing the numerical rational interpolation method described in the previous section determine $\Psi_g^{4,2(j)}\left(\bar{\theta}_{\mathbf{H}_g,h}^{(i)}, d, p^{(j)}\right)$ for each $h = 1,2,\ldots,\max(h)$[45]. Thus, given the matrix in step 4 the $\bar{q}$ pairs of points $\left(\theta_{\mathbf{H}_g,q}^{(i)}, RSS_g^{(j)}\left(\theta_{\mathbf{H}_g}^{(i)}, d, p^{(j)}\right)\right)$ required by the interpolation method are easily extracted since the needed RSS function values are given as columns.

6. The optimal location, $\widetilde{\bar{\theta}}_{\mathbf{H}_g}^{(i)}$, is given by:

$$\widetilde{\bar{\theta}}_{\mathbf{H}_g}^{(i)} = \underset{1\le p^{(j)}\le L}{\arg\min}\, AIC_g^{(j)}\left(\theta_{\mathbf{H}_g}^{(i)}, d, p^{(j)}\right)$$

where

$$AIC_g^{(j)}\left(\theta_{\mathbf{H}_g}^{(i)}, d, p^{(j)}\right) = N_g^{*(j)} \ln\left(\min_{1\le h\le \max(h)} RSS_g^{(j)}\left(\theta_{\mathbf{H}_g}^{(i)}, d, p^{(j)}\right) / N_g^{*(j)}\right) + 2p^{(j)}$$

for $1 \le p^{(j)} \le L$, while the best fit lag-order $\widetilde{p}^{(j)} = \underset{1\le p^{(j)}\le L}{\arg\min}\, AIC_g^{(j)}\left(\widetilde{\bar{\theta}}_{\mathbf{H}_g}^{(i)}, d, p^{(j)}\right)$ and we let $\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g} = \left[\theta_{\mathbf{H}_g}^{(1)}, \theta_{\mathbf{H}_g}^{(2)}, \ldots, \widetilde{\bar{\theta}}_{\mathbf{H}_g}^{(i)}, \ldots, \theta_{\mathbf{H}_g}^{(k-1)}\right]$.

7.

c. If $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right) = 1$:

For each $j \ne \bar{j}$, in turn, compute the updated $\mathbf{R}_g^{(j)}$-factor using rank-ones Givens updates (see §4.6.4)[46] and letting i signify which threshold the process in regime j is dependent on, it is fitting to construct the following matrix:

[45] Note that by setting $\max(h)$ as a very large number one is able to assess the value of the RSS function, via the interpolated function, at any point on the line defined by the set of points $\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)} \in \mathbf{H}_g$.

[46] Clearly, this step is only necessary if any of the regimes $j \ne \bar{j}$ are explicit functions of the threshold vector.

$$\begin{bmatrix} AIC_g^{(1)}\left(\theta_{\mathbf{H}_g}^{(1)}, d \mid p^{(1)}=1\right) & AIC_g^{(1)}\left(\theta_{\mathbf{H}_g}^{(1)}, d \mid p^{(1)}=2\right) & \cdots & AIC_g^{(1)}\left(\theta_{\mathbf{H}_g}^{(1)}, d \mid p^{(1)}=L\right) \\ AIC_g^{(2)}\left(\theta_{\mathbf{H}_g}^{(2)}, d \mid p^{(2)}=1\right) & \ddots & & \vdots \\ \vdots & & & \\ AIC_g^{(j)}\left(\tilde{\theta}_{\mathbf{H}_g}^{(j)}, d, \tilde{p}^{(j)}\right) & AIC_g^{(j)}\left(\tilde{\tilde{\theta}}_{\mathbf{H}_g}^{(j)}, d, \tilde{p}^{(j)}\right) & & AIC_g^{(j)}\left(\tilde{\tilde{\theta}}_{\mathbf{H}_g}^{(j)}, d, \tilde{p}^{(j)}\right) \\ \vdots & \ddots & & \vdots \\ AIC_g^{(k)}\left(\theta_{\mathbf{H}_g}^{(k)}, d \mid p^{(k)}=1\right) & \cdots & & AIC_g^{(k)}\left(\theta_{\mathbf{H}_g}^{(k)}, d \mid p^{(k)}=L\right) \end{bmatrix} \tag{4.12}$$

since then $\tilde{p}^{(j)} = \arg\min_{1 \le p^{(j)} \le L} AIC_g^{(j)}\left(\theta_{\mathbf{H}_g}^{(j)}, d, p^{(j)}\right)$, for $1 \le j \le k$, is easily extracted from the matrix. Also, $AIC_g^{(1,2,\ldots,k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \tilde{\mathbf{p}}_{\mathbf{H}_g}\right) = \sum_{j=1}^{k}\left(\min_{1 \le p^{(j)} \le L} AIC_g^{(j)}\left(\theta_{\mathbf{H}_g}^{(j)}, d, p^{(j)}\right)\right)$, where $\tilde{\mathbf{p}}_{\mathbf{H}_g} = \left[\tilde{p}^{(1)}, \tilde{p}^{(2)}, \ldots, \tilde{p}^{(j)}, \ldots, \tilde{p}^{(k)}\right]$, is simply the summation of the minimum AIC value in each row of the previously suggested matrix. Hence, the desired parameter values associated with $\mathbf{H}_g$ are given by $\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}$, $\tilde{\mathbf{p}}_{\mathbf{H}_g}$ and $AIC_g^{(1,2,\ldots,k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \tilde{\mathbf{p}}_{\mathbf{H}_g}\right)$.

d. If $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right) > 1$:

For each $j \neq \bar{j}$, in turn, compute the updated $\mathbf{R}_g^{(j)}$-factor using rank-ones Givens updates (see §4.6.4)[47] and for each of these regimes compute values of the RSS function for $1 \le p^{(j)} \le L$ using Givens corrections as explicated in §4.6.7. These values are conveniently exposited in a matrix with the following structure[48]:

$$\begin{bmatrix} 1 & 2 & \cdots & L \\ \left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g} \mid p^{(j)}=1\right) & \left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g} \mid p^{(j)}=2\right) & \cdots & \left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g} \mid p^{(j)}=L\right) \\ RSS_g^{(1)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(1)}=1\right) & RSS_g^{(1)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(1)}=2\right) & \cdots & RSS_g^{(1)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(1)}=L\right) \\ RSS_g^{(2)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(2)}=1\right) & \ddots & & \vdots \\ \vdots & & & \\ RSS_g^{(j)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(j)}=1\right) & RSS_g^{(j)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(j)}=2\right) & & RSS_g^{(j)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(j)}=L\right) \\ \vdots & & & \\ RSS_g^{(k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(k)}=1\right) & \cdots & & RSS_g^{(k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(k)}=1\right) \end{bmatrix} \tag{4.13}$$

where the first row indicates lag-order, the $k-1$ subsequent rows constitute the threshold vector that corresponds with a minimization of the RSS function for the

[47] Clearly, this step is only necessary if any of the regimes $j \neq \bar{j}$ are explicit functions of the threshold vector.

[48] Note that this structure is particularly convenient in the context of the, still to be detailed, proposed optimization routine.

lag-order stipulated in the first row and the $k \times L$ submatrix of RSS function values where it is clear that $RSS_g^{(j)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d \mid p^{(j)}\right) \equiv \Psi_g^{4,2(j)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}^{(i)}, d \mid p^{(j)}\right)$. Since, $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right) > 1$ implies that the hypercube of interest defines a subspace of the continuous threshold space that requires a search for an optimal point, this matrix will be returned to the optimization routine.

The optimization routine, detailed below, relies, to some extent, on the observation that $\Psi_g^{4,2(j)}\left(\boldsymbol{\theta}_{\mathbf{H}_g}^{(i)}, d \mid p^{(j)}\right)$ is particularly well-behaved.

Figure 1 illustrates typical shapes of the function defined over different hypercubes. As a result, it is pertinent, in light of our concern with computational expense, that we develop an optimization routine that optimally utilizes this characteristic. The routine that we propose will locate the optimal threshold point for $1 \le p^{(j)} \le L$ and $j = 1,2,\ldots,k$.

b) Multidimensional Interpolation-based Optimization

Two matrices and their interaction are central to the optimization routine that we outline in this section. The first, defined previously, is matrix (4.13) while the second is a matrix of L horizontally concatenated $\left(2 \times \left((k-1)\left(rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\Theta}\right)\right)\right)\right)$ sub-matrices containing the coordinates of the line-end points, which will define the domain of the L rational functions which we intend to interpolate. This matrix may be written as:

$$\left[V_1 \mid V_2 \mid \cdots \mid V_L\right] \tag{4.14}$$

where

$$V_l = \left[\begin{array}{cccc|cccc|c|cccc} \vec{\theta}_{\mathbf{H}_g,1}^{(1)} & \theta_{\mathbf{H}_g,1}^{(2)} & \cdots & \theta_{\mathbf{H}_g,1}^{(k-1)} & \theta_{\mathbf{H}_g,1}^{(1)} & \vec{\theta}_{\mathbf{H}_g,1}^{(2)} & \cdots & \theta_{\mathbf{H}_g,1}^{(k-1)} & \cdots & \theta_{\mathbf{H}_g,1}^{(1)} & \theta_{\mathbf{H}_g,1}^{(2)} & \cdots & \vec{\theta}_{\mathbf{H}_g,1}^{(k-1)} \\ \vec{\theta}_{\mathbf{H}_g,\infty}^{(1)} & \theta_{\mathbf{H}_g,1}^{(2)} & \cdots & \theta_{\mathbf{H}_g,1}^{(k-1)} & \theta_{\mathbf{H}_g,1}^{(1)} & \vec{\theta}_{\mathbf{H}_g,\infty}^{(1)} & \cdots & \theta_{\mathbf{H}_g,1}^{(k-1)} & \cdots & \theta_{\mathbf{H}_g,1}^{(1)} & \theta_{\mathbf{H}_g,1}^{(1)} & \cdots & \vec{\theta}_{\mathbf{H}_g,\infty}^{(k-1)} \end{array}\right]$$

for $l = 1,2,\ldots,L$. The multidimensional interpolation-based optimization routine that we purport may then be stipulated as:

1. Initialize matrices (4.13) and (4.14) by setting $\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g} \mid p^{(j)} = l\right) = \boldsymbol{\theta}_g$, for $l = 1,2,\ldots,L$, in (4.13) and the all of the submatrices in (4.14) equal to the line-end points that define the set of inclusive boundaries (see §4.5).
2. Let $j = 0,1,\ldots,\max(j)$, sequentially, indicate the number of iterations performed by the routine, where $\max(j)$ is the user-defined value assigned to the maximum allowable number of iterations.
    a. Determine the direction of interpolation *i*, defined as: $i = j \bmod(k-1)+1$.
    b. Group the set of optimal points ($\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g} \mid p^{(j)} = l\right)$, for $l = 1,2,\ldots,L$, in matrix (4.13) into subgroups so that all equivalent optimal points are contained within a single subgroup.
    c. Given the direction of interpolation *i*, determined in a., construct a matrix defining the line end-point coordinates. The number of line end-point coordinates is equivalent to the number of sub-groups determined in b.
    d. Consider the line end-point coordinates of each sub-group, in turn, and compute the value of the RSS function for each regime, given a particular lag order in that regime (as exposited in matrix 4.13), by using these end-point coordinates as inputs in the interpolation algorithm, described previously, as well as $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\mathbf{e}}\right)$.
    e. If the computed values for the RSS function, in d., are smaller than those of the previous iteration, for all $1 \le p^{(j)} \le L$, then the columns of matrix (4.13) that relate to the sub-group, determined in b. should be replaced by the corresponding values at the (more) optimal point and after updating matrix (4.14) for the new set of optimal points we perform the next iteration, by letting $j = j+1$ and performing a. through e. until the set of optimal points remains unchanged[49].
    f.
        i. If the set of optimal points defined in matrix (4.13), given the last iteration, are equivalent (implying that the optimal point

[49] It makes sense to stop the iterations of the optimization routine by setting a maximum for the number of allowable iterations, here denoted as $\max(j)$, as well as, setting a minimum for the decrease in the value of the RSS function. The latter, may indeed be necessary as a result of the use of floating point operators in the optimization routine.

$\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g} = \left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g} \mid p^{(j)} = l\right)$, for any $l = 1,2,\ldots,L$) – the optimal AIC value may be easily deduced by substituting each value of RSS value submatrix in (4.13) by its corresponding AIC value.

ii. If the set of optimal points differ across the different lag order candidates in matrix (4.13) then values for the AIC function are determined for each of these optimal points, that is we replicate matrix (4.13) for each optimal point, compute the statistics stipulated in i., above, and choose the optimal point, $\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}$, as the coordinate that corresponds with a minimization of the minimum AIC function.

g. The optimal point, $\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}$, the value of the AIC value computed as $AIC_g^{(1,2,\ldots,k)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \widetilde{\mathbf{p}}_{\mathbf{H}_g}\right) = \sum_{j=1}^{k} \min_{1 \le p^{(j)} \le L} AIC_g^{(j)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, p^{(j)}\right)$ and the associated best-fit lag order, given as $\widetilde{\mathbf{p}}_{\mathbf{H}_g} = \left[\widetilde{p}^{(k)}, \widetilde{p}^{(k-1)}, \ldots, \widetilde{p}^{(1)}\right]$, where $\widetilde{p}^{(j)} = \arg\min_{1 \le p^{(j)} \le L} AIC_g^{(j)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, p^{(j)}\right)$ for $j = 1,2,\ldots,k$, are then easily deduced using the matrix computed in f. and are subsequently returned as the output of the optimization routine being the optimal set of parameter values for the active threshold hypercube $\mathbf{H}_g$.

The 2-d contour graphs in Figure 1 illustrate how the, aforementioned, routine locates the set of optimal points in the L rational RSS functions. Letting the direction of interpolation, i, equal unity the first set of interpolations indicate that the optimal location along $\Psi_g^{4,2}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}^{(1)}, d, p^{(j)}\right)$ is at $\boldsymbol{\theta}_{\mathbf{H}_g,\max(h)}^{(1)}$ for $l = 1,2,\ldots,4$. Updating the domains of the RSS functions to be interpolated we locate the set of optimal points when interpolating $\Psi_g^{4,2}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}^{(2)}, d, p^{(j)}\right)$ which are equal to $\boldsymbol{\theta}_{\mathbf{H}_g,1}^{(2)}$ for $l = 1,2$ and 3 and $\boldsymbol{\theta}_{\mathbf{H}_g,\max(h)}^{(2)}$ for $l = 4$. Hence, given step f ii. of the proposed optimization routine both $\boldsymbol{\theta}_{\mathbf{H}_g,1}^{(2)}$ and $\boldsymbol{\theta}_{\mathbf{H}_g,\max(h)}^{(2)}$ would be used as plausible optimal points subject to further investigation, as stipulated above. Clearly, the parameter estimates that correspond with the optimal location in $\mathbf{H}_g$ will be carried forth as potential optimal estimates for the model as a whole.

4.6.7. Updating QR Factorizations: $\mathbf{p}_i \rightarrow \mathbf{p}_{i>1}$

By computing the Cholesky factor of the QR decomposition of each of the k $N_g^{*(j)} \times (L+1)$ augmented regressor matrices, $\left[\mathbf{F}_g^{(j)} \mid \Delta\mathbf{z}_g^{(j)}\right]$, given the set maximal dimensionality of the lag space, L, all desired terms may be determined. Expanding on the approach stipulated in Schlittgen (1997), consider the set of k QR factorizations of the augmented regressor matrices of each regime conditional on some candidate threshold vector $\boldsymbol{\theta}_g \in \mathbf{H}_g$

$$\left[\mathbf{F}_g^{(j)} \mid \Delta\mathbf{z}_g^{(j)}\right] = \left[\mathbf{Q}_g^{(j)} \begin{pmatrix} \mathbf{R}_{g,1}^{(j)} & \mathbf{d}_{g,1}^{(j)} \\ 0 & d_{g,2}^{(j)} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\right] \tag{4.15}$$

where, initially, $\mathbf{p}_i = \mathbf{\iota}_k L$ and $RSS_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right) = \sum_{j=1}^{k} \left\| d_{g,2}^{(j)} \right\|_2^2$. Update $\mathbf{p}_i$ by allowing $p^{(j)} = L-1$ for some plausible j, denoted $\mathbf{p}_i^*$, then the new (augmented) regressor matrix, of regime j, is equal to $\left[\mathbf{F}_g^{(j)} \mid \Delta\mathbf{z}_g^{(j)}\right]$ with the last column of the submatrix $\mathbf{F}_g^{(j)}$ deleted and $RSS_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i^*\right) = \left(\mathbf{d}_{g,1}^{(j)}(L)\right)^2 + \sum_{j=1}^{k} \left\| d_{g,2}^{(j)} \right\|_2^2$, where $\mathbf{d}_{g,1}^{(j)}(L)$ is the Lth element of the L-vector $\mathbf{d}_{g,1}^{(j)}$. Generalizing the expression to account for any lag order combination, $\mathbf{p}_i$, when $p^{(j)} < L$ for all j, results in the following expression

$$RSS_g^{(1,2,\ldots,k)}\left(\boldsymbol{\theta}_g, d, \mathbf{p}_i\right) = \sum_{j=1}^{k} \left(\left\| d_{g,2}^{(j)} \right\|_2^2 + \sum_{l=1}^{L-p^{(j)}} \left(\mathbf{d}_{g,1}^{(j)}\left(p^{(j)}+l\right)\right)^2 \right)$$

where $p^{(j)}$ denotes the lag order in regime j, as defined by $\mathbf{p}_i$, and $\mathbf{d}_{g,1}^{(j)}\left(p^{(j)}+l\right)$ is the $\left(p^{(j)}+l\right)th$ element of the L-vector $\mathbf{d}_{g,1}^{(j)}$. Consequently, values for the RSS function may be efficiently computed for each plausible lag order combination conditional on $\boldsymbol{\theta}_g \in \mathbf{H}_g$ and

d [50]. The, previously, exposited concepts and constructs are integrated into the algorithm that we propose in the next section.

4.7. The Proposed Procedure

Algorithmically, the discussion detailed previously may be implemented as follows:

1. Fix the set of user-defined parameters $\{k, D, L, \max(h), \pi_0, \pi_1, m\}$ where k is the number of regimes; D is the maximum feasible value of the threshold delay; L is the maximum feasible value of the AR lag orders; $\max(h)$ is the number of evaluation points that will be used when interpolating the RSS function; π_0 and π_1 are the lower and upper trimming parameters, respectively; and m is the minimum number of cases[51] required to exist in each regime.
2. Let $d = D$
   a. Arrange the data in arranged autoregression form i.e. construct the $(N^* \times (L+2))$ matrix $[\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots, \mathbf{z}_{t-L} \mid \Delta\mathbf{z}_t \mid \mathbf{z}_{t-d}]$ and arrange each case/row vector such that the vector of delay variable values is arranged in ascending order.
   b. Define the vector of candidate thresholds as $\mathbf{t} = [z_{(\tau_0)}, z_{(\tau_0+1)}, \ldots, z_{(\tau_1)}]'$ where $\tau_0 = [N^* \pi_0]$, $\tau_1 = [N^*(1-\pi_1)]$, where $z_{(i)}$ is the *ith* order statistic of the last column of the matrix in arranged autoregression form discussed in (a). Store this vector of candidate thresholds as a global variable.
   c. Assign a label to each of the candidate thresholds such that the label i is assigned to $t_{(i)}$, the *ith* order statistic of $\mathbf{t}$, and construct a vector of candidate threshold indices, equal in dimension to $\mathbf{t}$.
   d. Construct the grid $\mathbf{I}^{\Theta}$, based on the, previously constructed, vector of candidate threshold indices, by implementing the algorithm delineated in §4.4. Denote each row vector $\overline{\boldsymbol{\theta}}_g$, for $g = 1, \ldots, G$, where G is the number of row vectors in $\mathbf{I}^{\Theta}$.

[50] As mentioned by both Schlittgen (1997) and CFP the approach delivers a suboptimal solution to the least squares problem since $\boldsymbol{\iota}_i'(\boldsymbol{\iota}_i L - \mathbf{p}_i')$ cases are not utilized. It is argued, however, that since ultimately the TAR model parameters are estimated using all available information the computational gains of the algorithm more than compensates for the loss in estimation accuracy.

[51] In the sense of Tsay (1989).

e. For $g = 1,2,\ldots,G$, in order:

- Map $\bar{\boldsymbol{\theta}}_g$, back to the continuous threshold space Θ i.e. letting $\bar{\theta}_g^{(i)}$ denote the *ith* element of $\bar{\boldsymbol{\theta}}_g$, then the row vector with elements $\mathbf{t}_{(\bar{\theta}_g^{(i)})}$, for $i = 1,2,\ldots,k-1$, will define some support point in $\boldsymbol{\Theta}$, recalling that $\mathbf{t}$ was stored as a global variable in b..
- Partition the matrix described in a. into k sub-matrices by evaluating the value of the delay variable relative to the candidate threshold variable, of the previous step. Drop the last column of each of the k sub-matrices resulting in the (augmented) basis matrices $\left[\mathbf{F0}_g^{(j)} \mid \boldsymbol{\Delta z}_g^{(j)}\right]$.
- Compute k QR factorizations of the (augmented) basis matrices.
  1. If $g = 1$: Initialize the procedure by computing the orthogonal matrices $Q_1^{(j)}$ and the upper-triangular matrices $R_1^{(j)}$ for $j = 1,2,\ldots,k$. Store these matrices as global variables.
  2. If $g > 1$: Update the set of QR factorizations by identifying the case(s) that will be transferred from one regime to another (contiguous) regime; updating the affected QR factorizations by implementing the Givens updates stipulated in §4.6.2; substituting the QR factorizations stored in memory as global variables.
- Construct $\mathbf{H}_g$ as described in § 4.5, and make a decision as to the maximum number of iterations that the optimization routine will be allowed to make:
  1. If $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\boldsymbol{\Theta}}\right) = 0$:
  2. If $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\boldsymbol{\Theta}}\right) = 1$: Utilize the line interpolation algorithm delineated in § 4.4.6. to locate the optimal threshold location in $\mathbf{H}_g$, carrying forth $\widetilde{\boldsymbol{\theta}}_g$, $\widetilde{\mathbf{p}}_{\mathbf{H}_g}$ and $AIC_g^{(1,2,\ldots,k)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \widetilde{\mathbf{p}}_{\mathbf{H}_g}\right)$.
  3. If $rows\left(\overline{\mathbf{H}}_g \cap \mathbf{I}^{\boldsymbol{\Theta}}\right) > 1$: Utilize the optimization routine delineated in § 4.4.6. to locate the optimal threshold location in $\mathbf{H}_g$, carrying forth $\widetilde{\boldsymbol{\theta}}_g$, $\widetilde{\mathbf{p}}_{\mathbf{H}_g}$ and $AIC_g^{(1,2,\ldots,k)}\left(\widetilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \widetilde{\mathbf{p}}_{\mathbf{H}_g}\right)$.

- Minimize the sequence $AIC_g^{(1,2,\ldots,k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \tilde{\mathbf{p}}_{\mathbf{H}_g}\right)$, conditional on d, across threshold hypercubes i.e. $\left(\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{p}}\right) = \underset{1 \le g \le G}{\arg\min}\, AIC_g^{(1,2,\ldots,k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \tilde{\mathbf{p}}_{\mathbf{H}_g}\right)$

3. Let $d = d - 1$ and repeat step 2 until $d = 1$.
4. Following Tong and Lim (1980), the set of associated minimal Normalized AIC statistics are given by: $NAIC\left(\tilde{\boldsymbol{\theta}}, d, \tilde{\mathbf{p}}\right) = \min_{1 \le g \le G} AIC_g^{(1,2,\ldots,k)}\left(\tilde{\boldsymbol{\theta}}_{\mathbf{H}_g}, d, \tilde{\mathbf{p}}_{\mathbf{H}_g}\right) \Big/ N^*$.
5. Finally, using $\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{\mathbf{p}}\right) = \underset{1 \le d \le D}{\arg\min}\, NAIC\left(\tilde{\boldsymbol{\theta}}, d, \tilde{\mathbf{p}}\right)$ and all available information estimate $\boldsymbol{\Phi}^{(j)}$ for $j = 1,2,\ldots,k$ by OLS.

We investigate the performance of this procedure relative to a set of alternative fitting approaches in the next section.

4.8. Simulation Analysis

4.8.1. The Data Generating Mechanisms

Below we specify two instances of multiple-threshold SETAR models[52], presented in "first-difference" form, the parameters of which we purport may be efficiently estimated by the proposed procedure.

The Asymmetric Band-SETAR model (K3)

The symmetric Band-SETAR specification, on which the CFP procedure is based, is easily generalized to allow for asymmetries in both the speed of reversion to the inner-band, as well as, the dispersion of the thresholds around the unconditional mean of the process[53]. This model may be written as

[52] All models are assumed to be "globally" stationary in the sense of Balke and Fomby (1997).
[53] The conditional mean of the processes discussed in this paper is set to zero since only demeaned processes are considered.

$$\Delta z_t = \left[\sum_{i=1}^{p^{(1)}} \phi_i^{(1)}\left(z_{t-i} + \theta^{(1)}\right)\right] I\left(z_{t-d} < \theta^{(1)}\right) + \left[\phi_0^{(2)} + \sum_{i=1}^{p^{(2)}} \phi_i z_{t-i}\right] I\left(\theta^{(1)} \le z_{t-d} < \theta^{(2)}\right) + \left[\sum_{i=1}^{p^{(3)}} \phi_i^{(3)}\left(z_{t-i} - \theta^{(2)}\right)\right] I\left(z_{t-d} \ge \theta^{(2)}\right) + \varepsilon_t \tag{4.16}$$

A Likelihood Ratio test, due to Hansen (1997), may be employed to test whether restricting (4.16) to being symmetric significantly alters the maximized likelihood.

The Composite Band-SETAR model (K4)

The model may be written as

$$\Delta z_t = \left[\sum_{i=1}^{p^{(1)}} \phi_i^{(1)}\left(z_{t-i} + \theta^{(1)}\right)\right] I\left(z_{t-d} < \theta^{(1)}\right) + \left[\phi_0^{(j)} + \sum_{i=1}^{p^{(j)}} \phi_i^{(j)} z_{t-i}\right] I\left(\theta^{(j-1)} \le z_{t-d} < \theta^{(j)}\right) + \left[\sum_{i=1}^{p^{(4)}} \phi_i^{(4)}\left(z_{t-i} - \theta^{(3)}\right)\right] I\left(z_{t-d} \ge \theta^{(3)}\right) + \varepsilon_t \tag{4.17}$$

where $j = 2,3$. This model exhibits Band-SETAR type behavior when in an outer-regime (i.e. the outer thresholds serve as attractors) and once in an inner-regime the process reverts to the unconditional mean.

The following section details the experimental setup utilized in the Monte Carlo simulations, the results of which have been reported in the succeeding section.

4.8.2. Experimental Design

We consider models (4.16) and (4.17) as data generating mechanisms given the parameter settings stipulated in Table 1. The pseudo-randomly generated error process is $NID\left(0, \sigma_\varepsilon^2\right)$ with $\sigma_\varepsilon^2 = \{0.2, 0.4, 0.9\}$. The generated SETAR processes are given a 200 observation start-up period which is then corrected for. Each experiment is replicated 100 times. In the trend set by CFP, the simulation analysis will focus on the accuracy of the threshold estimates and their computational expense[54]. As such, the following measures will be employed

[54] All simulations were programmed in Ox 3.20 and run on a 2.8 GHz Pentium IV with 512 MB of RAM.

Table I
Data Generating Process Parameter Settings

K3 and K4 correspond with the Asymmetric- and Composite Band-SETAR processes detailed in models (4.16) and (4.17) respectively, while $\mathbf{p}$, d, $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$ denote the vector of lag-orders, the delay value, the threshold vector and the set of autoregressive coefficients in order.

DGP	$\mathbf{p}$	d	$\boldsymbol{\theta}$	$\boldsymbol{\Phi}$
K3	$p^{(1)}=1;$	$d=2$	$\theta^{(1)}=-0.50;$	$\phi_1^{(1)}=0.20;$
	$p^{(2)}=1;$		$\theta^{(2)}=0.30.$	$\phi_0^{(2)}=-0.35, \phi_1^{(2)}=1;$
	$p^{(3)}=1.$			$\phi_1^{(3)}=0.30.$
K4	$p^{(1)}=2;$	$d=1$	$\theta^{(1)}=-0.30;$	$\phi_1^{(1)}=0.10,\ \phi_2^{(1)}=0.40;$
	$p^{(2)}=1;$		$\theta^{(2)}=0.15;$	$\phi_0^{(2)}=0.30, \phi_1^{(2)}=-1.00;$
	$p^{(3)}=1;$		$\theta^{(3)}=0.50.$	$\phi_0^{(3)}=0.00, \phi_1^{(3)}=-0.3;$
	$p^{(4)}=2.$			$\phi_1^{(4)}=0.35,\ \phi_2^{(4)}=-0.10.$

- The Mean Bias of $\hat{\theta}^{(i)}$: $B_{\mu}=\sum\left(\hat{\theta}^{(i)}-\theta^{(i)}\right)/M;$
- The Median Bias of $\hat{\theta}^{(i)}$: $B_{\tau}=median\left(\hat{\theta}^{(i)}-\theta^{(i)}\right);$
- The Root Mean Squared Error of $\hat{\theta}^{(i)}$: $RMSE=\sqrt{\sum\left(\hat{\theta}^{(i)}-\theta^{(i)}\right)^2}\Big/M;$
- The Mean Absolute Deviation of $\hat{\theta}^{(i)}$: $MAD=median\left|\hat{\theta}^{(i)}-\theta^{(i)}\right|;$
- The Sample Variance of $\hat{\theta}^{(i)}$: $\sigma^2_{\hat{\theta}^{(i)}}=\mathrm{var}\left(\hat{\theta}^{(i)}\right);$
- The Average Root Mean Squared Error: $RMSE_{\mu}=\dfrac{\sum_{i=1}^{k-1}\sqrt{\left(\hat{\theta}^{(i)}-\theta^{(i)}\right)^2/M}}{(k-1)};$
- The Average Mean Absolute Deviation: $MAD_{\mu}=median\left(\dfrac{\sum_{i=1}^{k-1}\left|\left(\hat{\theta}^{(i)}-\theta^{(i)}\right)\right|}{(k-1)}\right);$
- The Average Sample Variance: $\sigma^2_{\hat{\theta}}=\sum_{i=1}^{k-1}\mathrm{var}\left(\hat{\theta}^{(i)}\right)/k-1;$
- The Mean computation time: $t_{\mu}=\sum\hat{t}/M;$
- The Median computation time: $t_{\tau}=median\left(\hat{t}/M\right).$

Where $\hat{t}$ is computation time, measured in minutes, and M denotes the number of simulations performed in a single Monte Carlo simulation. The following user-defined settings were used: $\max(h)=14$, $(\pi_0,\pi_1)=(0.15, 0.85)$, $\kappa=0.15$, $L=4$ and $D=2$ for the

experiment with K3 as the DGP, while it is set to unity for K4. All experiments are based on the grid and hypercube construction algorithms of §4.4 and §4.5.

We investigate the capacities of the following five fitting approaches:

1. The continuous fitting approach, discussed throughout this chapter (**F1**);
2. A Grid Search using both QR factorizations and Givens updates (**F2**);
3. A Grid Search using only QR factorizations (without Givens updates) (**F3**);
4. A Grid Search based on the conventional SLS fitting approach (**F4**);
5. The continuous fitting approach based on SCLS estimation (**F5**).

An important detail is that each of the five fitting approaches will evaluate the RSS function at *exactly* the same set of grid points. This is due to the choice of setting $\max(h)=14$, i.e. each activated hypercube, of dimension greater than one, defines 14^d grid points at which the RSS function is evaluated, where *d*, here, denotes the dimensionality of the non-overlapping hypercube. The reason for doing so is that we are then able to directly compare the, aforementioned, statistics over the various fitting approaches. The results are tabulated and discussed in the next sub-section.

4.8.3. Monte Carlo Simulation Results

Table II, expositing the results for the asymmetric Band-SETAR process (K3), shows that F1 is superior to F4: $\sigma^2_{\hat{\theta}^{(1)}}$ being between 41 and 60 percent that of F4; the RMSE of F1 falls between 71% and 94% that of F4; the MAD of F1, correspondingly, falls between 50% and 89% that of F4; the mean and median biases, generally, point in the same direction while their magnitudes differ. Curiously, F1 takes approximately 10% longer to estimate; this finding is particularly interesting since, F1 entails *k-1* interpolations, given a *k-1* dimensional threshold hypercube, in the context of the optimization routine, implying that a total of *14(k-1)* grid points are evaluated within a particular threshold hypercube, while F4 computes the value of the RSS function at 14^{k-1} points. The bias measures of F1 hint at the presence of a small degree of systematic bias, this pattern is similar for F4 however indicating that this may be a result of having computed the statistics on a relatively small sample.

Clearly, F2 and F3 are inferior to both F1 and F4 in terms of estimation accuracy as well as estimation time; as such they are uninteresting given our focus on computational

Table II
Simulation Results for the Asymmetric (k3) Band-SETAR Process

The statistics presented in this table have been defined in §4.8.2. 100 simulations are performed based on a grid of the type detailed in §4.4 We set $\pi_0 = \pi_1 = \kappa = 0.15$; the maximum number of allowable iterations for the continuous approach to *K-1*; and the minimum incremental improvement in the RSS function value of a regime at 1×10^{-5}.

σ_ε^2	$\theta^{(1)}$					$\theta^{(2)}$					$RMSE_\mu$	MAD_μ	t_μ	t_τ
	$\sigma^2_{\beta^{(1)}}$	B_μ	B_τ	*RMSE*	*MAD*	$\sigma^2_{\beta^{(2)}}$	B_μ	B_τ	*RMSE*	*MAD*				
F1: The Continuous Approach														
0.2	0.123	0.345	0.293	0.492	0.154	0.133	-0.086	-0.185	0.375	0.144	0.434	0.149	3.055	3.024
0.4	0.178	0.308	0.305	0.522	0.161	0.214	-0.059	-0.140	0.466	0.142	0.494	0.152	2.937	2.937
0.9	0.430	0.213	0.303	0.690	0.205	0.372	0.048	0.064	0.612	0.165	0.651	0.185	2.872	2.875
F2: GS - QR Factorizations with Givens Updating														
0.2	0.178	0.248	0.229	0.489	0.145	0.153	-0.116	-0.192	0.408	0.141	0.449	0.143	13.602	13.531
0.4	0.251	0.322	0.303	0.595	0.171	0.240	-0.010	-0.127	0.490	0.146	0.543	0.159	13.616	13.533
0.9	0.459	0.198	0.170	0.706	0.215	0.415	0.088	0.049	0.650	0.196	0.678	0.206	13.547	13.515
F3: GS - QR Factorizations without Givens Updating														
0.2	0.388	0.132	0.024	0.637	0.257	0.371	0.019	0.086	0.610	0.287	0.624	0.272	9.723	9.712
0.4	0.607	0.097	-0.087	0.785	0.356	0.598	0.096	0.172	0.779	0.337	0.782	0.347	9.723	9.720
0.9	1.309	0.011	-0.216	1.144	0.488	1.248	0.273	0.446	1.150	0.514	1.147	0.501	9.727	9.727
F4: GS - Conventional SCLS														
0.2	0.265	0.089	0.154	0.523	0.173	0.266	-0.189	-0.151	0.550	0.178	0.512	0.176	2.728	2.725
0.4	0.440	0.032	0.005	0.664	0.237	0.396	-0.145	-0.141	0.646	0.252	0.655	0.245	2.723	2.719
0.9	0.717	0.076	0.042	0.850	0.306	0.732	-0.042	-0.227	0.857	0.330	0.854	0.318	2.723	2.720
F5: Continuous Fitting Approach based on SCLS														
0.2	0.291	0.118	0.202	0.552	0.189	0.258	-0.164	-0.124	0.534	0.178	0.543	0.184	1.191	1.191
0.4	0.427	0.043	0.099	0.655	0.229	0.388	-0.160	-0.120	0.643	0.240	0.649	0.235	1.140	1.111
0.9	0.837	0.087	0.093	0.919	0.323	0.850	0.011	-0.170	0.922	0.330	0.921	0.327	1.118	1.111

efficiency[55]. They do however indicate that, for three-regime models, computing $\Delta \mathbf{z}_g^{(j)} - \mathbf{F}_g^{(j)}\left(\mathbf{F}_g^{(j)\prime}\mathbf{F}_g^{(j)}\right)^{-1}\left(\mathbf{F}_g^{(j)\prime}\Delta \mathbf{z}_g^{(j)}\right)$ is faster than updating $\mathbf{Q}_g^{(j)}$ and $\mathbf{R}_g^{(j)}$ via Givens transformations and that the optimization routine contributes significantly to both estimation speed and accuracy. Consequentially, we introduced a composite fitting approach (F5) which, when a particular hypercube has been activated, will compute the RSS function values via SCLS and use the interpolation-based optimization routine to locate the optimal point. The computational gains are significant requiring approximately 60% less time than either F1 or F4, while the estimation accuracy measures are very similar compared to that of F4.

Figure 2, which graphs estimation time against sample size for the asymmetric three-regime Band-SETAR model, illustrates that the computational gains of utilizing F5 increases considerably, relative to F1, as sample size increases, being approximately 90% faster when $N = 200$; the result remaining stable when comparing F5 with F4. These findings are particularly relevant if one is interested in conducting Monte Carlo experiments which require

[55] Clearly, computing the number of flops required by each algorithm would be optimal in a comparative investigation such as this. The complexity and length of the algorithms render the approach infeasible however. Instead, by way of utilizing an object-oriented programming approach we attempt to minimize dependence of the results on cross-program differences in programming efficiency.

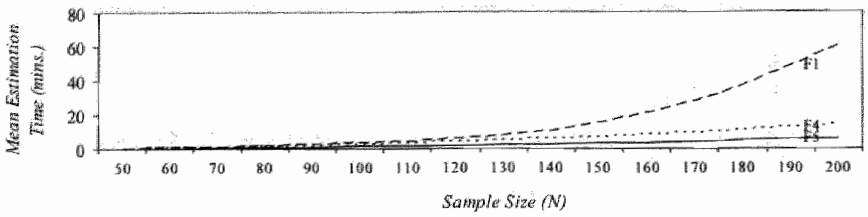

Figure 2. Mean Estimation Time vs. Sample Size in Asymmetric Three-Regime SETAR Models. The graphs plot the mean estimation times required by F1, F4 and F5 when the sample size is allowed to range from 50 through to 200. 25 simulations were performed with user-defined parameter settings equal to those used in the construction of Table II, except that σ_ε^2 is set to 0.4.

the estimation of models of this type, as we do in subsequent chapters, since, assuming that 1000 replications are made given a sample size of 200, one experiment would require approximately 1000 hours using F1, 240 hours using F4, and 100 hours using F5. Hence, the benefit of utilizing F5 when $k = 3$ is self-evident.

Table III presents the results of the simulations based on model (4.17). In this case F5 is superior in terms of both estimation time and the estimation bias measures; requiring approximately 74% of the time needed by F1 and 3% of that needed by F4, while being 10% to 20% more accurate as measured by the mean RMSE and MAD measures. Additionally, F5 generally outperforms F1 and F4 in terms of threshold estimate bias and variance. Here, as opposed to the three-regime case, however, the computational advantage of using Givens updates becomes apparent.

Figure 3 graphically displays the effect of sample size on estimation time when fitting four-regime Composite SETAR models. The results indicate congruence with our theoretic expectations since the advantages of using QR factorizations and Givens updates becomes more apparent. The results are clearly driven by the interpolation-based optimization routine developed in §4.6.6.

An important observation is that estimation time for F1 increases by a factor of 3.161 when shifting from fitting a three-regime as apposed to a four-regime model; the same figure

Table III
Simulation Results for the Composite (k4) Band-SETAR Processes

The statistics presented in this table have been defined in §4.8.2. 100 simulations are performed based on a grid of the type detailed in §4.4 We set $\pi_0 = \pi_1 = \kappa = 0.15$; the maximum number of allowable iterations for the continuous approach to $K-1$; and the minimum incremental improvement in the RSS function value of a regime at 1×10^{-5}.

σ_ε^2	$\theta^{(1)}$			$\theta^{(2)}$			$\theta^{(3)}$			$\sigma_{\hat{\varepsilon}}^2$	$RMSE_\mu$	MAD_μ	t_μ	t_τ
	B_{μ}	*RMSE*	*MAD*	B_{μ}	*RMSE*	*MAD*	B_{μ}	*RMSE*	*MAD*					
F1: The Continuous Approach														
0.2	-0.011	0.367	0.276	-0.082	0.376	0.297	-0.138	0.404	0.350	0.135	0.382	0.308	9.227	9.182
0.4	-0.178	0.591	0.512	-0.105	0.537	0.449	-0.045	0.518	0.455	0.318	0.549	0.472	9.284	9.250
0.9	-0.327	0.838	0.547	-0.021	0.726	0.632	0.141	0.710	0.579	0.595	0.758	0.586	9.260	9.200
F2: GS - QR Factorizations with Givens Updating														
0.2	-0.096	0.391	0.305	-0.165	0.409	0.297	-0.197	0.428	0.350	0.144	0.409	0.317	36.173	36.109
0.4	-0.114	0.559	0.449	-0.073	0.547	0.466	-0.038	0.526	0.475	0.299	0.544	0.463	36.393	36.181
0.9	-0.373	0.795	0.573	-0.060	0.732	0.656	0.128	0.711	0.599	0.493	0.746	0.609	36.275	36.128
F3: GS - QR Factorizations without Givens Updating														
0.2	-0.073	0.400	0.328	-0.140	0.416	0.335	-0.060	0.413	0.329	0.160	0.410	0.331	652.681	642.360
0.4	-0.134	0.674	0.587	-0.059	0.625	0.535	0.189	0.661	0.573	0.379	0.653	0.565	663.506	647.760
0.9	-0.380	1.016	0.792	-0.026	0.880	0.703	0.377	0.927	0.726	0.835	0.941	0.740	657.863	658.135
F4: GS - Conventional SCLS														
0.2	-0.063	0.404	0.297	-0.177	0.420	0.327	-0.175	0.442	0.380	0.160	0.422	0.335	212.927	212.028
0.4	-0.056	0.636	0.477	-0.093	0.638	0.442	0.011	0.620	0.504	0.401	0.631	0.474	208.364	206.435
0.9	-0.103	0.846	0.634	-0.081	0.861	0.553	0.049	0.812	0.635	0.597	0.839	0.607	208.364	206.435
F5: Continuous Fitting Approach based on SCLS														
0.2	0.033	0.378	0.273	-0.112	0.373	0.231	-0.125	0.420	0.323	0.142	0.390	0.276	6.867	6.758
0.4	-0.078	0.585	0.482	-0.084	0.557	0.409	0.020	0.550	0.436	0.336	0.564	0.442	6.832	6.716
0.9	-0.090	0.683	0.461	0.057	0.715	0.493	0.255	0.785	0.531	0.459	0.728	0.495	6.849	6.756

being 78.196 for F4 and 5.993 for F5, respectively. Table IV presents a set of conservative estimates of the time needed to estimate five- and six-regime SETAR models,[56] when $N = 100$. These estimates are simple linear extrapolations of the times required to estimate K3 and K4. They are conservative since we know that the number of candidate thresholds within each hypercube increases exponentially when increasing the dimensionality of the threshold space as apposed to a proportionate increase for both F1 and F5. The contribution of the proposed approach is apparent: being 570 and 14 000 times faster than the conventionally adopted SCLS approach when estimating five- and six-regime SETAR models, respectively[57].

In conclusion, it is clear that F5 is most appropriate when estimating three- and four-regime SETAR models, conditional on the functional forms of the models describing the processes in these regimes being explicitly dependent on the threshold vector, for all sample sizes. When the number of regimes in a SETAR model exceeds four, however, it is apparent that F1 is exceedingly superior from the perspective of computational expense and estimation

[56] Clearly, performing a comparative study using Monte Carlo simulations based on five and six-regime SETAR models is computationally intractable; for a six-regime model 100 simulations utilizing F4 would require approximately 250 years to compute. This is obviously largely attributable to the particularly dense grid that we ultimately construct: each hypercube having a 7529536 point grid embedded in it.

[57] This result would be exacerbated if more than two regimes are defined by processes that are explicit functions of the threshold vector.

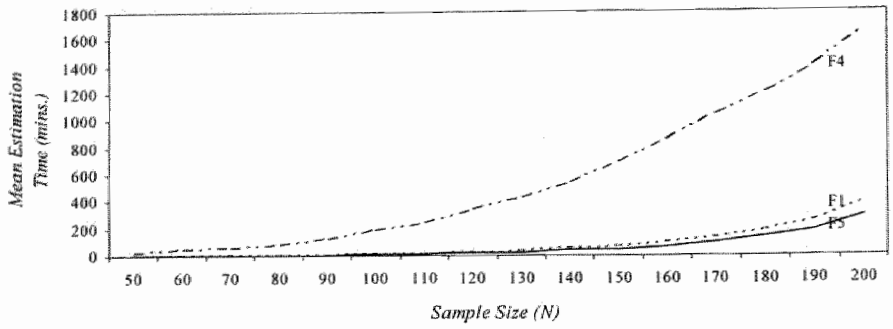

Figure 3. Mean Estimation Time vs. Sample Size in Composite Four-Regime SETAR Models. The graphs plot the mean estimation times required by F1, F4 and F5 when the sample size is allowed to range from 50 through to 200. 25 simulations were performed with user-defined parameter settings equal to those used in the construction of Table II, except that σ_ε^2 is set to 0.4.

accuracy; a conclusion that we draw from the simulation results based on K3 and K4. When a model is free of regimes that are explicitly dependent on the threshold vector, as is the case for model (4.2) it is clear that F4, which also most simply implemented from a programming perspective, is most appropriate.

4.9. Conclusion

This chapter proposes a generalization of the fitting approach advocated by CFP, applied to the estimation of multiple-threshold SETAR models. Hence, the generalization necessitates the construction of an appropriately restricted grid given requirements that are particular to SETAR models, as well as, defining a very specific path through the threshold space with the property that it is conducive to the implementation of updating algorithms. Based on the CFP result that the RSS function is continuous over non-overlapping threshold intervals we define the notion of non-overlapping threshold hypercubes. Additionally, we exposit how a grid may be constructed within a non-overlapping threshold hypercube. Given this foundation we discuss how QR factorizations and Givens updates of these factorizations may be used to compute values for the RSS function within a particular threshold subspace. Finally, we design a computationally expedient interpolation-based, numerical, optimization routine based on a series of rational interpolations of the RSS function. The simulation results suggested that combining the SCLS approach with the proposed optimization routine would yield favorable results. Investigating the length of time required to fit two examples of three-

Table IV
Linearly Extrapolating Mean Estimation Times of
Five and Six Regime SETAR Models

Mean estimation times, in minutes, given $N = 100$ whereby the reported, italicized, computation times for $k = 5$ and $k = 6$ are computed by linear extrapolation.

k	F1	F4	F5	Ratio $(F4/F1)$	Ratio $(F4/F5)$
3	2.937	2.723	1.140	0.927	2.389
4	9.284	212.028	6.832	24.020	32.641
5	*29*	*16500*	*41*	*570*	*400*
6	*93*	*1300000*	*245*	*14000*	*5300*

and four regime SETAR models indicated that the latter fitting approach is most expedient. Extrapolating the results to five and six regime models however poignantly illustrated the computational gains that can be achieved by utilizing the fitting approach that we advocate. Issues for further research relate to extending the technique to alternative TAR models, such as the M-TAR model introduced into the literature by Enders and Granger (1998), as well as, the possible application of grids with varying densities to the multiple-threshold SETAR model fitting problem.

5

Unit Root Testing in the Presence of Double-Threshold Nonlinearity

5.1. Introduction

It has been inveterately established that conventional unit root tests may suffer from a lack of power when the model under the alternative is not adequately veridical. Perron (1989) efficaciously demonstrated that Dickey-Fuller unit root tests may have little power when the true data-generating process is stationary around a broken linear trend; leading to the suggestion that the problem may be circumvented by the incorporation of dummy variables at the break date (see Perron (1989, 1993, 1994) and Perron and Vogelsang (1992, 1993)). Further complications were evinced by Kim *et al.* (2000) who showed that this solution may lead to spurious rejections of Perron's (1989) test. Paralleling these findings were observations in the nonlinear time-series literature where Pippenger and Goering (1993) and later Balke and Fomby (1997), in the context of threshold cointegration, found that the presence of a threshold boundary leads to a significant deterioration in the power of the Dickey-Fuller test; the effect

being amplified as the threshold effect becomes stronger. Ensuingly, unit root tests have been developed that specify a stationary threshold autoregression, or TAR model, under the alternative.

In this chapter we propose a unit root test specifying an unrestricted, stationary, three-regime M-TAR model under the alternative hypothesis; setting forth the general tendence initiated by the work of Enders and Granger (1998) who were the first to stipulate a threshold process under the alternative in a unit root test. Subsequent studies, amongst others by Berben and van Dijk (1999) and Kapetanios and Shin (2002), have shown that this testing strategy is worthwhile. The focal points of this study are momentum-TAR[58] variants of the Equilibrium- and Band-TAR models introduced into the literature by Balke and Fomby (1997). An important reason for doing so are the findings of Cook (2003) which indicate that practitioners will tend to detect M-TAR rather than TAR type nonlinearity irrespective of the nonlinearity that is actually present in an unknown process. Additionally, Enders and Granger's (1998) power test results indicate the superior performance of the M-TAR- over the SETAR-based test. Finally, Enders and Granger (1998) and Kapetanios and Shin (2002) have illustrated the superiority of three-regime TAR over two-regime TAR specifications under the alternative, while allowing a unit root process to govern the inner-regime. The performance of the proposed test is juxtaposed with that of conventional unit root tests in the form of the Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests, as well as, extant nonlinear unit root tests represented by the tests of Enders and Granger (1998) (henceforth EG) and Berben and van Dijk (1999) (henceforth BvD).

We begin by amalgamating two branches in the unit root testing literature in §5.2; those related to structural breaks defined in the time domain and those that allow for the presence of a threshold boundary in the levels of a process. §5.3 then details M-TAR model fitting approaches leading naturally to §5.4 which presents the direct unit root testing procedure that we espouse. Monte Carlo experiments are utilized in §5.5 to generate sets of critical values for the tests and presents a comprehensive assessment of its power and size properties. §5.6 details an empirical application of the tests in modeling the term structure of Eurocurrency interest rates while §5.7 concludes.

[58] Momentum-TAR, or M-TAR, models, have a lagged difference as the threshold variable (see Enders and Granger (1998)) as opposed to the better known SETAR model which has a lagged value as the threshold variable.

5.2. Globally Stationary Three-Regime M-TAR Processes

Suppose that z_t follows a trend stationary model which exhibits a break in period T_B, with $T_B = \lambda N$, where N denotes sample size and λ, typically referred to as the break fraction, pinpoints the time at which the change in the trend parameters occurs[59]. Consider the following two models, known as the changing growth model (5.1) and the model Montañés and Reyes (1998) refer to as the mixed model (5.2)

$$z_t = \mu_1 + \beta_1 t + (\beta_2 - \beta_1)(t - T_B)D_t + \xi_t \tag{5.1}$$

$$z_t = \mu_1 + (\mu_2 - \mu_1)D_t + \beta_1 t + (\beta_2 - \beta_1)tD_t + \xi_t \tag{5.2}$$

where t is a deterministic trend and D_t is a dummy variable being equal to unity when $t > T_B$ and is zero otherwise. Let

$$\xi_t = \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t \tag{5.3}$$

then analyzing the integration order of z_t may be achieved by computing the pseudo t-ratio of ϕ_1 in (5.3). Combining (5.1) and (5.3) results in the following convenient formulation

$$z_t = \begin{cases} \mu_1 + \beta_1 t + \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad t \le T_B \\ \mu_1 + \beta_1 t + (\beta_2 - \beta_1)(t - T_B) + \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad t > T_B \end{cases} \tag{5.4}$$

Now allowing for the possibility of m trend breaks[60], let $T_{B1}, T_{B2}, \ldots, T_{Bm}$ denote the time period of these breaks as a fraction of the sample size, and dividing this set of break points into two mutually exclusive subsets, "1" and "2", such that $z_{T_{Bj}-d} = \theta^{(2)}$ and $z_{T_{Bi}-d} = \theta^{(1)}$ for $i \neq j$ with $i, j \in \{1,2,\ldots,m\}$, we are in a position to reformulate (5.4) as

[59] Note that λN corresponds with a mapping of Berben and van Dijk's (1999) drifting threshold construct into the time domain.

[60] Additionally assuming that $z_t \sim I(0)$.

$$z_t = \mu_1 + \beta_1 t + \begin{cases} \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad z_{t-d} > \theta^{(2)} \\ (\beta_2 - \beta_1)(t - T_B) + \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad \theta^{(1)} \le z_{t-d} \le \theta^{(2)} \\ \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad z_{t-d} < \theta^{(1)} \end{cases} \tag{5.5}$$

where, since $(\beta_2 - \beta_1)(t - T_B) = z_{t-1} - z_{T_B - 1}$, we obtain

$$z_t = \mu_1 + \beta_1 t + \begin{cases} \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad z_{t-d} > \theta^{(2)} \\ \phi_0 + \sum_{i=1}^{p} \phi_i^{*} z_{t-i} + \varepsilon_t & \text{if} \quad \theta^{(1)} \le z_{t-d} \le \theta^{(2)} \\ \sum_{i=1}^{p} \phi_i z_{t-i} + \varepsilon_t & \text{if} \quad z_{t-d} < \theta^{(1)} \end{cases} \tag{5.6}$$

where $\phi_1^{*} = \phi_1 + 1$, $\phi_0 = -z_{T_B - 1}$ and "1" and "2", here, denote lower and upper regimes, respectively. Model (5.6) is a self-exciting threshold autoregressive, or SETAR, model; here defined around both a drift and a linear time trend[61]. Hence it is unsurprising that the findings stemming from the structural break literature are in congruence with those of the nonlinear time-series literature. Knowledge of this relation dates back to Quandt (1960) who showed that testing for linearity, utilizing SETAR type models, is algebraically quite similar to the issue of testing for a structural break of unknown timing.

Models, of which (5.6) is an example, were originally proposed by Tong (1978, 1983) and Tong and Lim (1980) as an alternative model for describing periodic time series. More generally a TAR model defines a time-series which may be represented as

$$z_t = \sum_{j=1}^{k} \left[\phi_0^{(j)} + \phi_1^{(j)} z_{t-1} + \ldots + \phi_{p^{(j)}}^{(j)} z_{t-p^{(j)}} \right] I_t\left(\theta^{(j-1)} \le v_{t-d} < \theta^{(j)}\right) + \varepsilon_t^{(j)} \tag{5.7}$$

where $d \in \Im^{+}$, $p^{(j)}$ is the autoregressive lag order in regime j, and $I_t(A)$ is an indicator function being equal to unity if event A occurs and is zero otherwise. The thresholds, $\left\{\theta^{(0)}, \ldots \theta^{(k)}\right\}$ where $-\theta^{(0)} = \theta^{(k)} = \infty$, partition the one-dimensional Euclidean space into k regimes such that the process follows a different linear AR model in each regime. Associated with each regime is the sequence of martingale difference sequences $\left\{\varepsilon_t^{(j)}\right\}$ satisfying $E\left[\varepsilon_t^{(j)} \mid F_{t-j}\right] = 0$, $\sup_t E\left[\left|\varepsilon_t^{(j)}\right|^{\delta} \mid F_{t-1}\right] < \infty$ a.s. for some $\delta > 2$ with F_{t-1} the σ - field generated

[61] The mixed model, (5.2), may be similarly reformulated.

by $\{\varepsilon_{t-i}^{(j)} \mid i = 1,2,\ldots; j = 1,\ldots,k\}$[62]. This process is non-linear if at least one non-trivial threshold separating two regimes, defined by different linear models, exists. For purposes of convenience model (5.7) is typically referred to as a TAR$(k;p;d)$ model where k denotes the number of regimes, p denotes the AR order and d is the threshold lag or delay parameter. Setting $v_{t-d} \equiv z_{t-d}$ results in the SETAR sub-class while allowing for the alternative setting where $v_{t-d} \equiv \Delta z_{t-d}$ results in another interesting sub-class of models, termed momentum, or M-TAR, models, by Enders and Granger (1998) who introduced them into the time-series literature.

Consider the following formulations of three-regime M-TAR models based on the Equilibrium and Band-TAR models, whose SETAR specifications have shown to be interesting models *per se*, after having been introduced into the literature by Balke and Fomby (1997) in the context of threshold cointegration[63,64,65]

$$z_t = \begin{cases} \sum_{i=1}^{p} \phi_i^{(1)} z_{t-i} + \varepsilon_t & if \quad \Delta z_{t-d} < \theta^{(1)} \\ z_{t-1} + \sum_{i=2}^{p} \phi_i^{(2)} z_{t-i} + \varepsilon_t & if \quad \theta^{(2)} > \Delta z_{t-d} > \theta^{(1)} \\ \sum_{i=1}^{p} \phi_i^{(3)} z_{t-i} + \varepsilon_t & if \quad \Delta z_{t-d} > \theta^{(2)} \end{cases} \tag{5.8}$$

$$z_t = \begin{cases} \sum_{i=1}^{p} \phi_i^{(1)} \left(z_{t-i} - \theta^{(1)}\right) + \varepsilon_t & if \quad \Delta z_{t-d} < \theta^{(1)} \\ z_{t-1} + \sum_{i=2}^{p} \phi_i^{(2)} z_{t-i} + \varepsilon_t & if \quad \theta^{(2)} > \Delta z_{t-d} > \theta^{(1)} \\ \sum_{i=1}^{p} \phi_i^{(3)} \left(z_{t-i} - \theta^{(2)}\right) + \varepsilon_t & if \quad \Delta z_{t-d} > \theta^{(2)} \end{cases} \tag{5.9}$$

The models describe processes that switch regimes dependent on the value of a past change relative to the threshold vector $\boldsymbol{\theta} = \left[\theta^{(1)}, \theta^{(2)}\right]$. Balke and Fomby introduced the more restrictive symmetric case for each model by setting $\theta^{(2)} = -\theta^{(1)} = \theta$, letting the inner-regime be a (driftless) random walk, and having equivalent outer-regime processes. Each process is propelled by a unit root when in the inner-regime and differs only with respect to how reversion takes place[66]. The Equilibrium-MTAR process, or EQ-MTAR process, (5.8), is so defined as to allow reversion to an equilibrium level while the Band-MTAR, (5.9), process

[62] Note that under special circumstances this need not hold for inner-regime processes.
[63] Throughout z_t is both a "demeaned" *and* "detrended" process.
[64] See §2.3.2 for the conditions that need to be satisfied to render the model stationary.
[65] Stationarity conditions for the "Returning-drift" or RD-TAR model which they also investigate are as yet undetermined. Consequently, we refrain from studying this model as a plausible stationary model under the alternative hypothesis in a unit root test.
[66] Conditional on the assumption of global stationarity.

reverts to an equilibrium-band; instead of a particular equilibrium level. Logically, each process has the property that it may be locally non-stationary while maintaining the global property of stationarity, under the assumption that the outer-regime processes are stationary. Subjecting model (5.7) to the requirement that the piecewise linear AR function be continuous everywhere results in a subclass of models which Chan and Tsay (1998) refer to as continuous TAR, or CTAR, models. It is easily shown that for $p = 1$ model (5.9) belongs to this subclass.

Figure 1 graphically depicts examples of EQ-TAR, EQ-MTAR, Band-TAR and Band-MTAR processes. Utilizing the cubic splines as a guide it is apparent that models (5.8) and (5.9) are capable of capturing dynamics of processes with structural breaks. The next section details M-TAR model parameter estimation after which we present the unit root test to which our interest in this chapter extends.

5.3. Fitting M-TAR Models

Least squares (LS) estimates are easily obtained if the threshold vector, the threshold delay, and the lag-order in each regime are known since then the model is linear in the remaining parameters. These estimates may be conveniently computed as

$$\hat{\mathbf{\Phi}}^{(j)} = \left(\mathbf{F0}^{(j)'}\mathbf{F0}^{(j)}\right)^{-1}\left(\mathbf{F0}^{(j)'}\mathbf{z}^{(j)}\right) \tag{5.10}$$

where $\mathbf{F0}^{(j)}$ and $\mathbf{z}^{(j)}$ are the matrix of regressor variables and the vector of the dependent variable in regime j respectively, for $j = 1,2,3$. Correspondingly, the residual variance may be computed as

$$\hat{\sigma}^2 = \sum_{j=1}^{k} \hat{\boldsymbol{\varepsilon}}^{(j)'}\hat{\boldsymbol{\varepsilon}}^{(j)}/N^* \tag{5.11}$$

where $\hat{\boldsymbol{\varepsilon}}^{(j)} = \mathbf{z}^{(j)} - \mathbf{F0}^{(j)}\hat{\mathbf{\Phi}}^{(j)}$ and $N^* = N - \max(d, p)$ is the effective sample size. Chan (1993) showed that the estimate in (5.10) is consistent and asymptotically Gaussian.

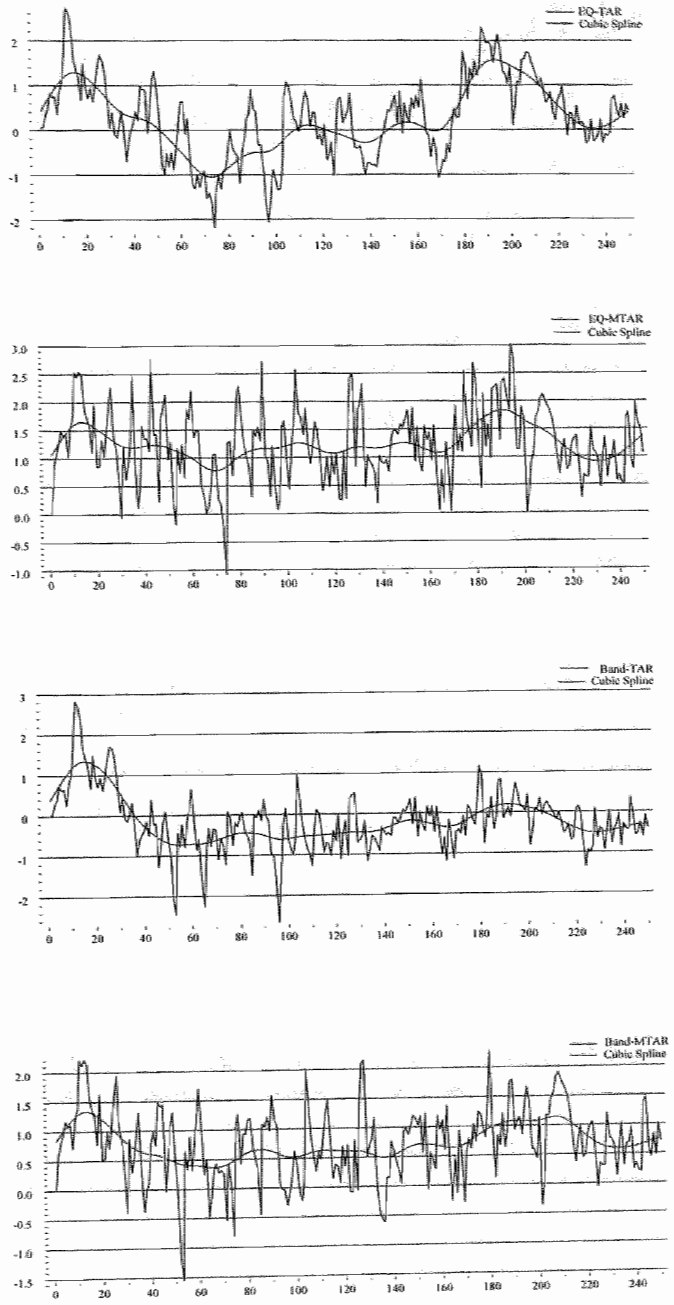

Figure 1. Simulated EQ-TAR, EQ-MTAR, Band-TAR and Band-MTAR processes.

Parameters settings: $N = 250$; $d = 3$; $\theta^{(1)} - 0.1$, $\theta^{(2)} = 0.9$ $\sigma_\varepsilon^2 = 0.2$; $\phi_0^{(1)} = -0.1$, $\phi_1^{(1)} = 0.2$; $\phi_0^{(2)} = 0$; $\phi_0^{(3)} = 0.9$, $\phi_1^{(3)} = 0.9$.

If, on the other hand, the threshold vector, $\boldsymbol{\theta}$, threshold delay, d, and the autoregressive lag-order, p, are unknown they too will need to be estimated. This may be achieved in a computationally tractable manner by way of employing the procedure advanced by Coakley *et al.* (2003), for the single threshold case, and the procedure that we developed in Chapter 4, for the multiple-threshold case. There we pedantically detailed a new fitting approach, based on a novel interpolation-based optimization routine, which is significantly less computationally expensive than the sequential conditional LS approach (SCLS), which is typically adopted, while achieving a similar degree of model fit[67].

Let $\boldsymbol{\Theta}$ define a $G \times 2$ matrix of candidate threshold vectors, with *gth* row vector $\boldsymbol{\theta}_g = \left[\theta_g^{(1)}, \theta_g^{(2)}\right]$; $\theta_g^{(1)}$ and $\theta_g^{(2)}$ meeting the requirement that they are elements of the set of order statistics of the delay variable. Further, each row vector must meet the requirements:

1. $\Delta z_{\left(\lfloor N^* \pi_0 \rfloor\right)} < \theta_g^{(1)} < \theta_g^{(2)} < \Delta z_{\left(\lfloor N^* (1-\pi_1) \rfloor\right)}$;
2. $N^* \kappa \leq N_g^{(j)*}$.

where $\Delta z_{(i)}$ denotes the *ith* order statistic of the delay variable; π_0 and π_1 are the upper and lower trimming parameters; $N^* = \sum_{j=1}^{3} N_g^{(j)*}$ is the effective sample size defined as the sum of the number of cases[68] in each regime; κ is an *a priori* restriction on the minimal fraction of observations that must be present in each regime; and $\lfloor \cdot \rfloor$ is a floor function here defined as the integer part operator.

If the functional form of the models defined over each regime in a particular TAR model are not explicit functions of the threshold vector, as is the case for model (5.8), then sequential conditional least squares estimation will correspond with maximum likelihood estimation. If, alternatively, at least one regime in a TAR model is explicitly dependent on the threshold vector, as is the case for model (5.9), the continuous threshold space Θ[69] will have to be searched for the optimal location of the thresholds. To this end, non-overlapping threshold hypercubes, so defined that they partition the 2-dimensional threshold space into subspaces

[67] Resultantly we utilize it in the simulations performed in subsequent sections. Note that in Chapter 4 we advocate two fitting approaches, the first, or (**F1**), is not suited to the task of estimating M-TAR models since it relies on a series of updating algorithms that are inapplicable to the current context, while the second approach (**F5**), SCLS in combination with the proposed multidimensional interpolation-based optimization routine, is easily malleated into a form capable of estimating M-TAR parameters.

[68] In the sense of Tsay (1989).

[69] Practically, the continuous threshold space is restricted by the aforementioned requirements.

over which the RSS function of a TAR model is continuous and rational[70], will need to be constructed[71]. Summarily, each non-overlapping threshold hypercube, $\mathbf{H}_g$, includes a single support point $\boldsymbol{\theta}_g \in \boldsymbol{\Theta}$ and covers the area defined by $\boldsymbol{\theta}_g$ and the set of contiguous grid points, $\boldsymbol{\theta}_{g^*}$, that fulfill the condition $\theta_g^{(i)} \le \theta_{g^*}^{(i)}$ for $i = 1,2$ [72].

Let D and L be *a priori* choices of the maximum allowable values of the threshold delay and autoregressive lag-order, respectively. Then, considering each non-overlapping threshold hypercube $\mathbf{H}_g$, for $g = 1,2,\ldots,G$, in turn, we aim to locate the optimal threshold location, within $\mathbf{H}_g$, which corresponds with a minimization of the RSS function conditional on some $1 \le p \le L$. Then $\widetilde{p}_g = \underset{1\le p\le L}{\arg\min}\, AIC\left(\widetilde{\theta}_g, d, p\right)$ is the optimal lag-order, correspondingly, $\widetilde{\boldsymbol{\theta}}_g = \underset{\boldsymbol{\theta}\in\mathbf{H}_g}{\arg\min}\left(RSS_g\left(\boldsymbol{\theta}, d, \widetilde{p}_g\right)\right)$ is the optimal location of the thresholds where $RSS\left(\boldsymbol{\theta}, d, \widetilde{p}_g\right) = \sum_{j=1}^{3}\left(\mathbf{z}_g^{(j)} - \mathbf{F}_g^{(j)}\hat{\boldsymbol{\Phi}}_g^{(j)}\right)'\left(\mathbf{z}_g^{(j)} - \mathbf{F}_g^{(j)}\hat{\boldsymbol{\Phi}}_g^{(j)}\right)$ with $\hat{\boldsymbol{\Phi}}_g^{(j)} = \left(\mathbf{F}_g^{(j)'}\mathbf{F}_g^{(j)}\right)^{-1}\left(\mathbf{F}_g^{(j)'}\mathbf{z}_g^{(j)}\right)$ where $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}$ is a first-degree polynomial matrix, for $j = 1,3$; $\mathbf{F0}_g^{(j)}$ is the $\left(N_g^{*(j)} \times \widetilde{p}_g\right)$ regressor matrix and

$$\mathbf{U}_g^{(j)} = \left[c, c, \ldots, c\right]'\left[\theta^{(i)}, \theta^{(i)}, \ldots, \theta^{(i)}\right] \tag{5.12}$$

is a matrix of equivalent order; for $j = 1,3$ $c = -1$ while for $j = 2$ $c = 0$. Also, $AIC_g\left(\widetilde{\boldsymbol{\theta}}_g, d, \widetilde{p}_g\right) = \min_{1\le p\le L} AIC\left(\widetilde{\boldsymbol{\theta}}_g, d, p\right)$ denotes Tong's (1983) Akaike information criterion measure where $AIC_g\left(\widetilde{\boldsymbol{\theta}}_g, d, p\right) = N^* \ln\left(RSS\left(\widetilde{\boldsymbol{\theta}}_g, d, p\right)/N^*\right) + 2kp$. Hence, for each $\mathbf{H}_g$ we have estimates of $\boldsymbol{\theta}$, p, and AIC; estimated conditionally on d. Then D sets of parameter estimates will result when allowing $\left(\widetilde{\theta}^*, \widetilde{p}^*\right) = \underset{1\le g\le G}{\arg\min}\, AIC_g\left(\widetilde{\boldsymbol{\theta}}_g, d, \widetilde{p}_g\right)$, leading to the final (LS) estimates

[70] As proven by Coakley *et al.* (2003).
[71] We refer readers to Chapter 4 for a pedantic exposition and discussion of this construct. There we generalize the notion of a non-overlapping threshold interval, as exposited in Coakley *et al.* (2003), to the multiple-threshold TAR model case.
[72] Strictly speaking, only boundaries of $\mathbf{H}_g$ that contain $\boldsymbol{\theta}_g$ are included in the threshold hypercube.

$$\left(\hat{\boldsymbol{\theta}}, \hat{p}, \hat{d}, \hat{\sigma}^2\right) = \underset{1 \le d \le D}{\arg\min}\, NAIC(d) \tag{5.13}$$

where $NAIC(d) = AIC\left(\tilde{\theta}^*, d, \tilde{p}^*\right)/N^*$ is the Normalized AIC suggested by Tong and Lim (1980).

Chan (1993), generalizing Petrucelli's (1986) result, showed that for arbitrary p the conditional LS estimates of the parameters of a TAR $(k; p; d)$ model are strongly consistent[73]; additionally showing that $\hat{\boldsymbol{\theta}}$ is asymptotically independent of $\hat{\boldsymbol{\Phi}}^{(j)}$, for all j, and that the asymptotic distribution of each $\hat{\boldsymbol{\Phi}}^{(j)}$ are equal to the case where the thresholds are known.

5.4 A Unit Root Test Specifying Stationary MTAR Adjustment under the Alternative

Assuming that all values of r that satisfy the inverse characteristic equation $1 - \sum_{i=1}^{p-1} \rho_i r^i = 0$ lie outside the unit circle, ensuring the presence of no more than a single unit root, models (5.8) and (5.9) may be conveniently formulated in "first-difference" form as

$$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)} z_{t-1} + \varepsilon_t & if \quad \Delta z_{t-d} < \theta^{(1)} \\ \gamma^{(3)} z_{t-1} + \varepsilon_t & if \quad \Delta z_{t-d} > \theta^{(2)} \end{cases} \tag{5.14}$$

$$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)}\left(z_{t-1} - \theta^{(1)}\right) + \varepsilon_t & if \quad \Delta z_{t-d} < \theta^{(1)} \\ \gamma^{(3)}\left(z_{t-1} - \theta^{(2)}\right) + \varepsilon_t & if \quad \Delta z_{t-d} > \theta^{(2)} \end{cases} \tag{5.15}$$

where $\gamma^{(j)} = \phi_1^{(j)} - 1$ and it is easily shown that $\sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} = \sum_{i=1}^{p-1} \rho_i \Delta\left(z_{t-i} - \theta\right)$. As opposed to the testing procedure of Enders and Granger (1998) we propose a direct test of the unit root hypothesis[74]; the null hypothesis of a unit root versus the alternative of threshold stationarity being given by

$H_0 : \gamma^{(1)} = \gamma^{(3)} = 0$ vs.

[73] This was proven for the single threshold case. It is generally accepted that these results apply to the multiple threshold case.

[74] A similar testing strategy was been adopted by Caner and Hansen (2001) and Kapetanios and Shin (2002).

$$H_1 : \gamma^{(1)} < 0 \quad and \quad \gamma^{(3)} < 0 \tag{5.16}$$

in models (5.14) and (5.15). Note that the formulations in (5.14) and (5.15) are threshold model equivalents of the Augmented Dickey-Fuller test. Under the auxiliary assumption that $\varepsilon_t \sim i.i.d.$ from the theories of Davies (1977, 1987) and Andrews and Ploberger (1994), a test with near optimal power against distant alternatives of H_0 is the standard F-statistic:

$$W = \frac{N^*}{2}\left(\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}\right) \tag{5.17}$$

where $\tilde{\sigma}^2$ is the estimated residual variance under the, previously stated, null hypothesis and $\hat{\sigma}^2$ is computed as is done in (5.13).

If $\boldsymbol{\theta}$ is unknown it is appropriate to utilize the supremum-Wald test statistic, given estimates $\hat{d}$ and $\hat{p}$, which may be formulated as

$$W^{\sup}(\hat{d}, \hat{p}) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} W(\boldsymbol{\theta}, \hat{d}, \hat{p}) \tag{5.18}$$

where

$$W(\boldsymbol{\theta}, \hat{d}, \hat{p}) = \frac{N^*}{2}\left(\frac{\tilde{\sigma}_n^2(\hat{p}) - \hat{\sigma}^2(\boldsymbol{\theta}, \hat{d}, \hat{p})}{\hat{\sigma}^2(\boldsymbol{\theta}, \hat{d}, \hat{p})}\right) \tag{5.19}$$

and $\hat{\sigma}^2(\boldsymbol{\theta}, \hat{d}, \hat{p})$ is the residual variance given some candidate threshold vector; defined such that the elements are equal to an order statistic of delay variable for models like (5.14). As alluded to previously, estimating the F-statistic for model (5.15) is more complex as it requires that each non-overlapping threshold hypercube[75] be searched by subjecting the RSS function, which is rational and continuous there within[76], to the standard optimality condition.

[75] This construct is extensively detailed in §4.5.
[76] The proof of this statement can be found in the appendix of Coakley *et al.* (2003).

Finally, it is convenient to denote the sup-Wald test statistics corresponding with models (5.14) and (5.15) by $W_{EQ}^{sup}\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right)$ and $W_{Band}^{sup}\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right)$, respectively. Then the proposed tests are given by

$$W_{Type}^{sup}\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right) = \frac{N'}{2}\left(\tilde{\sigma}^2(\hat{p}) - \hat{\sigma}^2\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right) / \hat{\sigma}^2\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right)\right) \tag{5.20}$$

where $Type \in \{EQ, Band\}$, $\hat{\sigma}^2\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right)$ is the residual variance, given the threshold estimate $\hat{\boldsymbol{\theta}}$, best-fit threshold delay and lag-order parameter estimates, $\hat{d}$ and $\hat{p}$, respectively, and $\tilde{\sigma}^2(\hat{p})$ is the residual variance, given $\hat{p}$ lagged changes, under the H_0. Kapetanios and Shin (2002) derive the asymptotic distribution of the Wald statistic, for both known and unknown values of the threshold parameters, when a SETAR model is specified under the alternative, while Berben and van Dijk (1999), using a drifting threshold construct to circumvent the problems of related to the erratic limiting behavior of the indicator functions, do the same for the CTAR model. A common finding is the invariance of the limiting distribution of the Wald statistic to the location of the threshold.

Since $\boldsymbol{\theta}$ is not identified under H_0 the asymptotic distribution of $W_{Type}^{sup}\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right)$ is non-standard (i.e. not χ^2), and since bootstrap methods, such as Hansen's (1996) fixed-regressor bootstrap, have been shown to improve finite sample results substantially (see Kapetanios and Shin (2002)), it is useful to employ these methods in the current context. The next section presents critical values and power and size tests of (5.20) relative to a battery of linear and nonlinear unit root tests.

5.5 Critical Values and Power Tests

Given the non-standard distribution of $W_{Type}^{sup}\left(\hat{\boldsymbol{\theta}}, \hat{d}, \hat{p}\right)$ we conduct a Monte Carlo experiment to generate the critical values necessary to test hypothesis (5.16). To this end we generate 2000 driftless random walk processes, i.e.

$$z_t = z_{t-1} + \varepsilon_t \qquad \text{for } t = 1, \ldots, N \tag{5.21}$$

Table I

Critical Values for Rejecting the Null Hypothesis of a Unit Root when $\hat{d} = 1$

Panel A presents critical values for $W_{EQ}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$ while Panel B presents critical values for $W_{Band}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$. Critical values are based on 2000 simulations.

Sample Size	Probability of a smaller value											
	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
	No lagged Changes			One lagged Change			Two lagged Changes			Three lagged Changes		
Panel A: *Three-Regime EQ-MTAR model specified under the Alternative Hypothesis*												
No Estimated Deterministic Components												
50	8.96	10.18	13.58	9.25	10.70	14.93	8.67	10.14	13.43	8.35	9.57	13.51
100	8.03	9.01	11.27	8.23	9.42	12.51	7.76	8.96	11.81	7.22	8.32	10.76
200	7.50	8.64	11.25	7.57	8.79	10.80	7.27	8.22	10.15	6.82	8.02	9.97
Estimated Constant Attractor												
50	9.06	10.54	13.17	9.49	10.89	14.31	8.83	10.15	13.46	8.30	9.64	13.60
100	8.51	9.73	12.72	8.64	10.08	12.48	8.03	9.31	11.92	7.72	8.99	10.90
200	8.18	9.09	10.81	8.50	9.84	11.81	7.91	8.96	10.91	7.46	8.64	11.00
Estimated Trend Attractor												
50	10.57	12.28	15.38	11.36	13.30	17.98	10.29	12.16	15.66	9.58	11.11	15.06
100	10.23	11.61	14.01	10.65	12.20	15.38	9.89	11.10	13.23	9.37	10.73	13.05
200	10.11	11.41	13.67	10.41	11.60	14.59	9.67	10.86	13.05	9.24	10.55	13.08
Panel B: *Three-Regime Band-MTAR model is specified under the Alternative Hypothesis*												
No Estimated Deterministic Components												
50	9.16	10.65	14.22	9.51	11.01	15.86	8.89	10.56	14.90	8.40	9.88	13.87
100	8.26	9.37	12.91	8.58	9.82	13.07	7.86	9.00	11.31	7.40	8.51	11.34
200	7.75	8.78	11.52	7.85	8.96	10.96	7.57	8.57	10.50	6.99	8.11	9.88
Estimated Constant Attractors												
50	9.38	10.74	13.82	9.75	11.50	15.78	9.16	10.55	14.54	8.53	9.85	13.32
100	8.69	10.07	12.60	9.08	10.53	13.32	8.38	9.70	12.43	8.00	9.24	11.32
200	8.47	9.52	12.17	8.70	10.05	12.28	8.20	9.30	11.16	7.73	8.79	10.83
Estimated Trend Attractors												
50	11.12	12.90	16.69	11.68	13.96	19.00	10.73	12.73	16.39	9.96	12.03	15.77
100	10.62	11.88	14.74	10.85	12.32	15.43	9.99	11.23	14.03	9.69	10.89	13.96
200	10.25	11.47	13.83	10.61	11.81	14.18	10.00	11.29	13.86	9.53	10.83	13.69

where the pseudo-randomly generated error process is drawn from a standard Normal distribution and $N = 50, 100, 200$. Each generated random walk process is assigned a 200 observation startup period which is subsequently corrected for and the initial value, i.e. z_0, is randomized. Ensuingly, models (5.14) and (5.15) are estimated, whereby we allow the number of lagged changes to vary from 0 through 3 and values of the threshold delay to vary from 1 through 3[77]. As is the case for the Augmented Dickey-Fuller Test there is no simple way to jointly test for the presence of deterministic components and a unit root. Resultantly, we compute three sets of critical values:

[77] We additionally set $\pi_0 = \pi_1 = \kappa = 0.15$.

Table II

Critical Values for Rejecting the Null Hypothesis of a Unit Root when $\hat{d} = 2$

Panel C presents critical values for $W_{EQ}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$ while Panel D presents critical values for $W_{Band}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$. Critical values are based on 2000 simulations.

Sample Size	Probability of a smaller value											
	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
	No lagged Changes			One lagged Change			Two lagged Changes			Three lagged Changes		
Panel C: *Three-Regime EQ-MTAR model specified under the Alternative Hypothesis*												
	No Estimated Deterministic Components											
50	8.62	10.23	12.40	9.15	10.75	12.93	8.85	10.30	13.81	8.03	9.51	11.92
100	8.01	9.15	11.58	8.31	9.68	12.58	7.94	8.96	12.68	7.30	8.49	11.00
200	7.70	8.63	11.12	7.96	8.87	10.91	7.40	8.47	10.41	6.90	8.01	9.89
	Estimated Constant Attractor											
50	9.02	10.58	13.56	9.40	11.12	15.61	8.68	10.15	13.04	8.22	9.75	13.07
100	8.54	9.78	12.68	8.92	10.40	13.39	8.30	9.74	12.37	7.84	9.37	11.57
200	8.29	9.35	11.83	8.70	9.92	12.37	7.96	9.32	11.53	7.41	8.68	10.81
	Estimated Trend Attractors											
50	10.99	12.59	16.30	11.77	13.42	18.41	10.83	12.25	15.92	9.86	11.64	15.71
100	10.33	11.94	14.45	10.65	12.39	15.18	9.89	11.25	14.23	9.47	11.05	13.35
200	10.02	11.33	13.49	10.29	11.54	13.60	9.84	10.91	13.14	9.24	10.36	12.48
Panel D: *Three-Regime Band-MTAR model is specified under the Alternative Hypothesis*												
	No Estimated Deterministic Components											
50	9.06	10.90	13.98	9.54	11.09	14.19	9.07	10.55	13.53	8.32	9.87	12.84
100	8.36	9.59	11.86	8.58	10.15	13.38	8.06	9.19	11.91	7.59	8.76	11.30
200	7.93	8.94	11.26	8.17	9.20	11.23	7.63	8.61	10.55	7.05	8.30	10.18
	Estimated Constant Attractor											
50	9.31	10.77	13.69	9.88	11.23	14.78	9.15	10.67	13.45	8.66	10.23	13.31
100	8.80	9.81	12.31	9.13	10.26	12.79	8.67	9.95	12.42	8.07	9.43	11.68
200	8.49	9.54	12.81	8.88	10.14	12.72	8.15	9.38	11.82	7.67	8.94	11.38
	Estimated Trend Attractors											
50	11.31	13.19	16.32	11.89	13.43	15.94	11.06	12.62	15.82	10.33	11.89	16.17
100	10.57	12.13	14.84	10.84	12.47	15.29	10.09	11.29	13.79	9.47	11.14	13.40
200	10.34	11.74	14.17	10.65	12.07	14.43	10.04	11.09	13.27	9.40	10.72	12.87

1. Without estimating deterministic components.
2. With the estimation of a constant attractor; achieved by computing $W_{Type}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$ on the residuals, $\dot{z}_t$, in a regression of z_t on a constant $\dot{z}_t = z_t - a_0$.
3. With the estimation of both a constant and trend attractors; achieved by computing $W_{Type}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$ on the residuals, $\dot{z}_t$, in $\dot{z}_t = z_t - a_0 - a_1 t$.

For $\hat{d} = 1$ the critical values, at the 10%, 5%, and 1% significance levels, for $W_{EQ}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$ and $W_{Band}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$ are reported in panels A and B of Table I, respectively. There, for example, $W_{EQ}^{sup}(\hat{\theta}, 1, 3)$, for $N = 100$, exceeds 9.31 in approximately 5% of the 2000 trials when not accounting for deterministic components. Tables II and III represent equivalent computations

Table III

Critical Values for Rejecting the Null Hypothesis of a Unit Root when $\hat{d} = 3$

Panel E presents critical values for $W^{sup}_{EQ}(\hat{\theta}, \hat{d}, \hat{p})$ while Panel F presents critical values for $W^{sup}_{Band}(\hat{\theta}, \hat{d}, \hat{p})$. Critical values are based on 2000 simulations.

Sample Size	Probability of a smaller value											
	90%	95%	99%	90%	95%	99%	90%	95%	99%	90%	95%	99%
	No lagged Changes			One lagged Change			Two lagged Changes			Three lagged Changes		
Panel E: *Three-Regime EQ-MTAR model specified under the Alternative Hypothesis*												
	No Estimated Deterministic Components											
50	9.28	10.88	14.03	9.61	11.22	15.19	8.86	10.60	14.67	8.11	9.40	12.36
100	8.18	9.37	11.72	8.51	9.81	12.41	7.67	8.83	11.19	7.20	8.46	10.73
200	7.73	8.78	11.31	8.02	9.07	11.34	7.45	8.41	10.75	6.91	8.01	9.91
	Estimated Constant Attractor											
50	9.25	10.79	13.10	9.69	11.04	15.12	9.00	10.52	14.82	8.23	9.57	13.29
100	8.58	9.81	12.61	9.10	10.29	13.42	8.28	9.50	11.86	7.85	9.04	11.63
200	8.20	9.48	11.45	8.55	9.65	12.17	8.01	9.17	11.94	7.41	8.54	11.11
	Estimated Trend Attractors											
50	11.01	12.87	17.68	11.89	14.12	19.11	10.98	12.84	17.02	10.19	11.55	15.79
100	10.25	11.63	14.71	10.70	12.17	16.05	9.93	11.43	14.06	9.41	10.73	13.46
200	10.03	11.30	13.46	10.38	11.78	14.11	9.63	10.94	13.05	9.06	10.37	12.64
Panel F: *Three-Regime Band-MTAR model is specified under the Alternative Hypothesis*												
	No Estimated Deterministic Components											
50	9.46	11.06	14.40	9.98	11.47	15.07	8.98	10.75	14.27	8.28	9.99	12.56
100	8.47	9.64	11.87	8.68	9.96	12.20	7.95	9.13	11.94	7.41	8.72	11.26
200	7.87	8.88	11.65	8.22	9.29	11.94	7.69	8.61	11.10	7.11	8.32	10.82
	Estimated Constant Attractor											
50	9.33	10.95	13.93	10.19	12.02	15.72	9.50	10.99	15.04	8.55	10.30	13.87
100	8.74	10.08	13.40	9.15	10.33	13.47	8.48	9.75	12.12	8.06	9.20	11.58
200	8.51	9.76	12.07	8.79	9.94	12.16	8.24	9.47	12.06	7.64	8.76	11.17
	Estimated Trend Attractors											
50	11.55	13.05	17.48	11.90	13.57	17.34	11.12	12.91	17.40	10.66	12.28	16.35
100	10.42	11.75	14.99	10.82	12.23	15.80	10.16	11.64	14.59	9.59	10.70	13.13
200	10.26	11.65	14.03	10.60	11.94	14.69	9.97	11.30	13.83	9.43	10.68	13.26

for $\hat{d} = 2$ and 3. Hence, if $W^{sup}_{Band}(\hat{\theta}, \hat{d}, \hat{p})$ is estimated for a given process, comprising, say, 200 observations, when accounting for the presence of trend attractors, and $(\hat{\theta}, \hat{d}, \hat{p}) = (\hat{\theta}, 3, 2)$ the test indicates a rejection of the unit root hypothesis if $W^{sup}_{Band} > 9.29$. Clearly, a conservative approach to using the test statistics is to estimate $W^{sup}_{Type}(\hat{\theta}, \hat{d}, \hat{p})$ under the assumption of the presence of a constant and a trend since rejecting the null of a unit root under this assumption will imply correctly rejecting the null when these components are not present in the process under consideration.

Tables IV and V present power statistics over various parameter constellations while allowing for linear, single-threshold and double threshold data generating processes. Each generated process is regressed on a constant; the choice of doing so is *ad hoc* but typical in the literature. Additionally, we let $D = 3$, $L = 4$ and require the presence of at least 15 cases[78] in each regime.

[78] In the sense of Tsay (1989).

Table IV
Power Tests when the DGP is a Three-Regime Threshold Process

The figures are the percentage of instances that the tests correctly reject the unit root hypothesis. Critical values for the ADF and PP tests were obtained from Hamilton (1994), those for the Enders-Granger tests were obtained from Enders (1999).Each generated, lag-augmented data generating process, is demeaned before the tests are conducted and we let $\phi_2^{(1)} = \phi_2^{(3)} = 0.1$ in all cases. For the nonlinear tests we set $L = 4$ and $\pi_0 = \pi_1 = \kappa = 0.15$. For the proposed tests we let $D = 3$. The Enders and Granger tests are based on a consistent estimate of the attractor. Results are based on 500 simulations.

$\theta^{(1)}$	$\theta^{(2)}$	$\phi_1^{(1)}$	$\phi_1^{(3)}$	Augmented Dickey-Fuller Test			Phillips-Perron Test			Enders-Granger Test (SETAR)			Enders-Granger Test (M-TAR)			Berben-van Dijk Test			$W_{EQ}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$			$W_{Bond}^{sup}(\hat{\theta}, \hat{d}, \hat{p})$		
				10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel A: DGP is a Three-Regime M-TAR model																								
Symmetric Threshold Dispersion																								
-0.5	0.5	0.9	0.9	13.2	7.4	2.0	6.0	3.2	2.2	12.2	6.4	1.2	7.4	4.2	0.6	18.2	12.6	6.2	23.6	11.0	2.8	23.8	12.8	4.6
-0.5	0.5	0.8	0.8	30.4	20.8	8.0	29.2	17.8	9.6	19.0	10.4	2.2	18.0	11.0	3.0	42.6	29.4	14.0	57.2	40.2	13.6	60.8	43.0	16.0
-0.5	0.5	0.6	0.6	81.4	67.2	37.6	95.8	88.8	75.2	68.0	51.6	26.2	71.0	57.8	33.2	94.8	88.8	72.8	99.0	94.8	81.8	100.0	94.8	82.6
-0.5	0.5	0.4	0.4	96.0	90.8	70.8	100.0	100.0	99.4	90.2	81.6	64.6	89.0	85.0	70.2	100.0	99.8	99.0	100.0	100.0	99.4	100.0	100.0	99.0
-0.5	0.5	0.2	0.2	98.8	97.2	84.4	100.0	100.0	100.0	95.0	90.0	79.2	94.4	91.8	82.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Symmetric Equilibrium Reversion																								
-3.0	3.0	0.2	0.2	10.4	5.8	1.2	3.2	1.2	0.0	15.0	9.0	3.0	8.6	6.2	2.2	20.0	15.0	8.0	28.0	17.2	9.0	30.6	18.8	11.6
-1.5	1.5	0.2	0.2	65.0	50.2	21.4	75.0	65.0	47.8	60.8	46.0	23.4	77.2	69.8	54.2	83.4	75.4	55.8	95.8	96.0	84.0	97.4	98.0	85.4
-1.0	1.0	0.2	0.2	91.4	85.0	58.8	99.4	98.4	95.8	88.0	79.8	63.2	93.2	90.8	79.6	99.2	99.2	97.0	100.0	99.8	99.8	100.0	100.0	100.0
-0.5	0.5	0.2	0.2	98.8	97.2	84.4	100.0	100.0	100.0	95.0	90.0	79.2	94.4	91.8	82.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Panel B: DGP is a Three-Regime SETAR model																								
Symmetric Threshold Dispersion																								
-0.5	0.5	0.9	0.9	14.0	7.2	1.6	6.2	4.0	1.8	10.6	5.4	1.00	7.2	4.2	0.6	21.4	14.0	6.4	26.0	11.6	2.4	27.6	13.0	4.2
-0.5	0.5	0.8	0.8	42.6	28.8	11.4	47.4	33.4	19.4	25.6	15.8	4.4	22.0	12.0	2.8	55.4	43.4	25.8	60.8	47.4	16.6	63.4	48.8	16.8
-0.5	0.5	0.6	0.6	93.6	86.2	57.2	100.0	99.4	96.8	81.8	70.0	38.8	70.2	58.8	33.8	99.6	99.4	94.8	98.6	95.4	74.6	100.0	95.4	74.8
-0.5	0.5	0.4	0.4	98.8	95.0	83.2	100.0	100.0	100.0	93.4	88.0	72.4	89.0	80.8	63.4	100.0	100.0	100.0	100.0	100.0	99.8	100.0	100.0	100.0
-0.5	0.5	0.2	0.2	99.8	98.2	90.0	100.0	100.0	100.0	97.0	95.0	83.4	94.0	89.8	78.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Symmetric Equilibrium Reversion																								
-3.0	3.0	0.2	0.2	84.4	72.6	37.6	97.8	93.2	74.2	82.6	73.4	48.4	72.2	61.6	35.0	99.0	97.4	86.2	99.2	97.4	81.6	100.0	96.6	75.0
-1.5	1.5	0.2	0.2	97.4	95.0	74.0	100.0	100.0	100.0	94.0	92.2	83.0	88.6	82.0	64.0	100.0	100.0	100.0	100.0	100.0	99.2	100.0	100.0	100.0
-1.0	1.0	0.2	0.2	98.8	96.2	84.2	100.0	100.0	100.0	97.4	94.4	86.8	92.6	87.2	76.2	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0	100.0
-0.5	0.5	0.2	0.2	99.8	98.2	90.0	100.0	100.0	100.0	97.0	95.0	83.4	94.0	89.8	78.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table V
Power Tests when the DGP is an AR or a Two-Regime Threshold Process

The figures are the percentage of instances that the tests correctly reject the unit root hypothesis. Critical values for the ADF and PP tests were obtained from Hamilton (1994), those for the Enders-Granger tests were obtained from Enders (1999).Each generated, lag-augmented data generating process, is demeaned before the tests are conducted and we let $\phi_2^{(1)} = \phi_2^{(3)} = 0.1$ in all cases. For the nonlinear tests we set $L = 4$ and $\pi_0 = \pi_1 = \kappa = 0.15$. For the proposed tests we let $D = 3$. The Enders and Granger tests are based on a consistent estimate of the attractor. Results are based on 500 simulations.

θ	$\phi_1^{(1)}$	$\phi_1^{(3)}$	Augmented Dickey-Fuller Test			Phillips-Perron Test			Enders-Granger Test (SETAR)			Enders-Granger Test (M-TAR)			Berben-van Dijk Test			$W_{EQ}^{sup}(\hat{\theta},\hat{d},\hat{p})$			$W_{Band}^{sup}(\hat{\theta},\hat{d},\hat{p})$		
			10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Panel C: DGP is a Two-Regime M-TAR model																							
Symmetric Thresholds																							
0.5	0.2	0.9	88.8	80.6	49.6	100.0	100.0	100.0	84.2	78.2	63.0	92.4	88.4	82.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	0.2	0.8	94.6	88.0	63.2	100.0	100.0	100.0	88.8	82.6	69.0	94.0	90.4	81.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	0.2	0.6	97.6	94.6	81.6	100.0	100.0	100.0	93.0	87.6	75.0	94.4	89.6	80.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	0.2	0.4	99.4	97.4	90.2	100.0	100.0	100.0	94.4	90.8	81.4	95.0	92.8	83.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	0.2	0.2	99.8	98.8	93.6	100.0	100.0	100.0	96.0	93.2	85.8	97.6	93.8	85.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Panel D: DGP is a Two-Regime SETAR model																							
Symmetric Thresholds																							
0.5	0.2	0.9	22.4	16.6	7.0	31.8	26.8	22.2	38.8	30.0	19.0	21.0	14.8	7.0	78.6	75.8	68.0	52.6	45.2	26.2	80.6	76.4	65.2
0.5	0.2	0.8	73.0	59.2	29.4	94.2	88.8	80.4	78.8	67.0	47.2	68.4	56.2	35.2	99.8	99.4	97.4	96.8	91.2	73.2	100.0	100.0	95.6
0.5	0.2	0.6	96.2	92.4	75.8	100.0	100.0	100.0	93.0	88.4	74.2	90.4	85.8	71.6	100.0	100.0	100.0	100.0	100.0	99.4	100.0	100.0	100.0
0.5	0.2	0.4	99.6	97.8	88.2	100.0	100.0	100.0	95.2	90.8	81.8	95.4	92.6	81.6	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0.5	0.2	0.2	99.8	98.8	93.6	100.0	100.0	100.0	96.0	93.2	85.8	97.6	93.8	85.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Panel E: DGP is an AR model																							
NA	NA	0.9	12.8	7.2	2.0	6.6	3.6	2.0	11.2	5.2	0.8	7.2	4.0	0.6	21.6	14.4	6.0	23.2	11.6	3.0	26.4	9.4	2.6
NA	NA	0.8	44.0	28.8	12.4	54.4	39.2	22.6	29.2	17.6	5.2	25.8	14.6	4.6	61.6	47.6	28.6	65.0	50.6	19.2	58.0	41.2	18.2
NA	NA	0.6	94.2	87.2	59.4	100.0	99.8	98.6	81.6	71.2	50.8	79.0	68.6	49.4	99.8	99.6	97.4	99.6	98.6	88.0	98.6	94.4	82.6
NA	NA	0.4	99.0	96.2	86.8	100.0	100.0	100.0	92.4	86.4	75.2	91.8	86.8	77.2	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
NA	NA	0.2	99.8	98.8	93.6	100.0	100.0	100.0	96.0	93.2	85.8	97.6	93.8	85.8	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

The power statistics presented in Panels A and B of Table IV are computed on processes that are generated by three-regime M-TAR and SETAR models. In comparing the ADF with the PP tests it is apparent that generally the ADF is dominated by the PP except when the outer-regime processes are near unit roots. As may be expected the EG-M-TAR outperforms the EG-SETAR test when the DGP is a three-regime M-TAR; a result that is reversed when the DGP is a three-regime SETAR process. Interestingly, the

BvD test has superior power to both the EG tests this does however come at the expense of a substantial size distortion (see Table VI). The $W_{Band}^{sup}\left(\hat{\theta},\hat{d},\hat{p}\right)$ outperforms $W_{EQ}^{sup}\left(\hat{\theta},\hat{d},\hat{p}\right)$ for all parameter settings in terms of power while being thwarted by similar size distortions; of a magnitude that are comparable to that of the BvD test.

The results confirm the findings of Enders and Granger (1998) that their test is outperformed by the ADF as well as the findings of Berben and van Dijk (1999) who show that their test outperforms the EG tests and the ADF. Interestingly, the PP test, which hasn't as yet been utilized in comparing power and size properties in the literature performs remarkably well: being slightly inferior to the BvD test in terms of power yet not having the sizable size distortion that we alluded to previously.

Table V presents a set of similar computations when the DGP is a single-threshold M-TAR, single-threshold SETAR or a linear AR model. There the pattern present in Table IV is repeated: The PP test, outperforming the ADF and EG tests, is inferior to the BvD test, which, in turn, outperforms the $W_{EQ}^{sup}\left(\hat{\theta},\hat{d},\hat{p}\right)$ test and is outperformed by the $W_{Band}^{sup}\left(\hat{\theta},\hat{d},\hat{p}\right)$ test. Hence, the $W_{Band}^{sup}\left(\hat{\theta},\hat{d},\hat{p}\right)$ test is superior to all alternatives given a wide range of data generating mechanisms but suffers from a sizable distortion in size. The latter not being true at the 1% significance level as is evident in Table VI.

5.6. M-TAR Stationarity: An Empirical Application using the Term Structure of Eurocurrency Interest Rates

A pervasive empirical regularity in the study of the term structure of interest rates is that short-term and long-term interest rates are cointegrated. Recently, the asymmetric nature of the

Table VI
Size Tests

The probability of rejecting the true null hypothesis of a unit root. Critical values for the ADF and PP tests were obtained from Hamilton (1994), those for the Enders-Granger tests were obtained from Enders (1999).Each generated, lag-augmented data generating process, is demeaned before the tests are conducted and we let $\phi_2^{(1)} = \phi_2^{(3)} = 0.1$ in all cases. For the nonlinear tests we set $L = 4$ and $\pi_0 = \pi_1 = \kappa = 0.15$. For the proposed tests we let $D = 3$. The Enders and Granger tests are based on a consistent estimate of the attractor. Results are based on 500 simulations.

	Size statistics		
	10%	5%	1%
Augmented Dickey-Fuller Test	13.4	7.4	1.6
Phillips-Perron Test	5.4	2.4	1.4
Enders-Granger Test (SETAR)	6.8	3.0	0.4
Enders-Granger Test (M-TAR)	6.8	3.0	0.4
Berben-van Dijk Test	17.0	10.2	5.0
$W_{EQ}^{sup}\left(\hat{\theta}, \hat{d}, \hat{p}\right)$	18.6	9.6	1.8
$W_{Band}^{sup}\left(\hat{\theta}, \hat{d}, \hat{p}\right)$	18.8	10.8	1.6

equilibrium reverting behavior of the interest rate differential has come to light[79] (see amongst others Enders and Granger (1998), Georgoutsos and Kouretas (2001) and Bohl and Siklos (2002)). By way of illustrating the appropriate application of the statistics developed in previous sections we investigate the interest rate differential between discrete quarterly observations, in percent per annum, of the 1 and 12 month U.K. Eurocurrency interest rates over the period 1976:1 through 1998:4. The data are from the Harrisbank Database[80]. Figure 2, graphically depicting these time-series observations, affirms the likely stationarity of the interest rate differential. Our interest extends towards testing whether equilibrium reversion of the differential is asymmetric and stationary, as well as, whether this asymmetry is best captured by a single- or a double-threshold model.

We estimate the following models on the interest rate differential time-series, denoted z_t:

ADF: $$\Delta z_t = \gamma^{(1)} z_{t-1} + \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \varepsilon_t \tag{5.22}$$

EG (SETAR): $$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)} z_{t-1} + \varepsilon_t & if \quad z_{t-d} < \theta \\ \gamma^{(2)} z_{t-1} + \varepsilon_t & if \quad z_{t-d} \geq \theta \end{cases} \tag{5.23}$$

[79] Note that these studies investigate the restrictive case where the cointegrating residual is governed by a two-regime SETAR/M-TAR model.
[80] The Harrisbank Database is maintained by Richard Levich to which we extend our gratitude for making the data available.

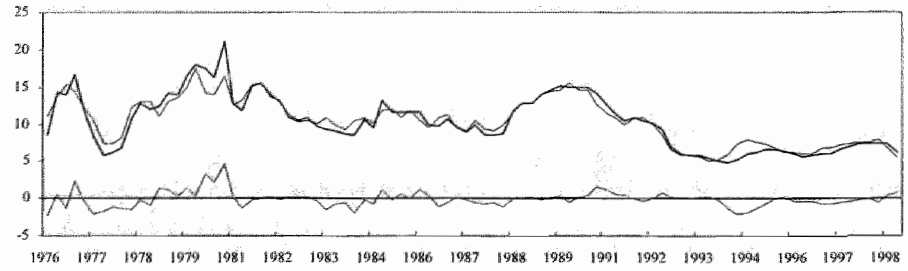

Figure 2. Short-term (1-month), long-term (12-month) and their differential for quarterly U.K. Eurocurrency interest rates. The dark solid line is the 1-month rate, the dashed line is the 12-month rate and the light solid line is the interest rate differential.

EG (M-TAR):
$$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)} z_{t-1} + \varepsilon_t & if \quad \Delta z_{t-d} < \theta \\ \gamma^{(2)} z_{t-1} + \varepsilon_t & if \quad \Delta z_{t-d} \geq \theta \end{cases} \tag{5.24}$$

BvD (CTAR):
$$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)} (z_{t-1} - \theta) + \varepsilon_t & if \quad \Delta z_{t-d} < \theta \\ \gamma^{(2)} (z_{t-1} - \theta) + \varepsilon_t & if \quad \Delta z_{t-d} \geq \theta \end{cases} \tag{5.25}$$

EQ-SETAR:
$$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)} z_{t-1} + \varepsilon_t & if \quad z_{t-d} < \theta^{(1)} \\ \gamma^{(3)} z_{t-1} + \varepsilon_t & if \quad z_{t-d} \geq \theta^{(2)} \end{cases} \tag{5.26}$$

Band-SETAR:
$$\Delta z_t = \sum_{i=1}^{p-1} \rho_i \Delta z_{t-i} + \begin{cases} \gamma^{(1)} (z_{t-1} - \theta^{(1)}) + \varepsilon_t & if \quad z_{t-d} < \theta^{(1)} \\ \gamma^{(3)} (z_{t-1} - \theta^{(2)}) + \varepsilon_t & if \quad z_{t-d} \geq \theta^{(2)} \end{cases} \tag{5.27}$$

Table VII presents statistical estimates of models (5.22) through (5.27), as well as, models (5.14) and (5.15). Our estimation strategy involves setting *a priori* constraints on the maximal values of the threshold delay and lag-order parameters, D and L, and using Tong's (1983) AIC measure to pinpoint the best-fit lag order while Tong and Lim's (1980) normalized AIC is used to estimate d. Additionally, we institute the requirement that at least 15% of the observations are in each regime.

In comparing the various models it is apparent that the Band-MTAR model is most apt at describing the interest rate differential. The double-threshold nonlinearity statistic clearly rejects the null of linearity, as represented by model (5.22), given the significance level of 0.004% computed using Hansen's (1996) fixed regressor bootstrap. Significance levels of the

linearity test for the other models clearly indicate that the nonlinearity is of the Band-TAR type; since the Band-TAR model is the only other model to indicate the rejection of the linearity hypothesis, be it at the 10% significance level.

An interesting observation is the positive, yet statistically insignificant, estimate of $\gamma^{(l)}$. Balke and Fomby (1997) discuss the possibility of a stationary TAR model, coined the returning-drift TAR model, that follows a unit root process in each regime yet reverts to equilibrium by way of the drift parameters. Hence, this observation does not annul the stationarity argument. Instead, there is strong evidence in favor of stationarity when considering the $W_{Band}^{sup}\left(\hat{\theta}, \hat{d}, \hat{p}\right)$ statistic; computed to be 36.409 it clearly exceeds the critical value of 12.06 (see Table III) at the 1% significance level. In fact, all tests reject the unit root hypothesis at the 5% level.

Based on the AIC[81], which we use to determine the best-fit lag order, the Band-MTAR model appears to fit the data more adequately than the alternatives. The double-threshold models have superior fit to the single threshold models, while they, in turn, outperform the linear model. The diagnostic portmanteau test of the residuals indicates that the first 12 autocorrelations are jointly insignificant for each of the models considered. The process tends to be governed by a driftless random walk when in the inner corridor, the threshold estimates being -0.688 and -0.250, and revert back to this inner-band when in an outer-regime; the strength of the reversion process being substantially stronger when the process passes through the lower threshold than when it does the upper. The Band-MTAR model (t-statistics are reported in parentheses) is therefore estimated as

$$\begin{aligned}\Delta z_t &= \underset{(1.12)}{0.109} + \underset{(0.28)}{0.032}\Delta z_{t-1} + \underset{(1.22)}{0.064}\Delta z_{t-2} + \\ &\underset{(0.81)}{0.172}(z_{t-1} + 0.688)I(\Delta z_{t-3} < -0.688) - \underset{(-5.42)}{0.714}(z_{t-1} + 0.25)I(\Delta z_{t-3} > -0.250)\end{aligned} \tag{5.28}$$

[81] Alternatively, the more parsimonious Schwartz Bayesian criterion (SBC) or the Hannan-Quinn criterion (HQC) may be used (see Kapetanios (1999)). The SBC, which we don't state explicitly in Table VII, yields the same best-fit lag order for the Band-MTAR model.

Table VII
Model Estimates of the Interest-Rate Differential for the U.K.

[a] The maximum allowable value of the threshold delay parameter D is set to 3. [b] The maximum allowable number of lagged changes is set to 3. [c] l and u indicate whether the coefficient estimate relates to the lower or upper regime, respectively. [d] Numbers between parentheses in this row are t-statistics for the null hypothesis $\rho^{(i)} = 0$. [e] The AIC is computed as $AIC = N^* \ln\left(RSS\left(\hat{\theta}, \hat{d}, \hat{p}\right)\right) + 2P$, with $P = (\hat{p} + 2)$ (excluding the ADF model for which $P = (\hat{p} + 1)$), where $N^* = N - \max\left(\hat{d}, \hat{p}\right)$ and $\hat{p}$ denotes the number of included lagged changes. [f] $W^{sup}_{Type}\left(\hat{d}, \hat{p}\right)$ is the supremum-Wald test statistic used in testing the unit root hypothesis. [g] The supremum-Wald test statistic used in testing the single-threshold nonlinearity and double-threshold nonlinearity hypothesis (see Hansen 1999). [h] The entries in this row are significance levels of the nonlinearity hypothesis computed using Hansen's (1996, 1999) fixed-regressor bootstrap method (2000 replications). [i] The Ljung-Box statistic testing whether the first 12 residual autocorrelations are jointly equal to 0. Significance levels are in parentheses.

		Single-Threshold Models			Double-Threshold Models			
	ADF Model	EG SETAR Model	EG MTAR Model	BvD CTAR Model	EQ-TAR Model	Band-TAR Model	EQ-MTAR Model	Band-MTAR Model
$\hat{d}$ [a]	NA	3	3	3	3	3	3	3
$\hat{p}$ [b]	3	3	3	3	0	2	0	2
$\gamma^{(l)}$ [c]	-0.434 (-3.591) [d]	-0.688 (-7.152)	-0.666 (-7.200)	-0.728 (-7.507)	-0.612 (-3.913)	-0.588 (-3.660)	0.173 (0.811)	0.172 (0.812)
$\gamma^{(u)}$	NA	-0.226 (-2.639)	0.141 (1.707)	-0.306 (-3.541)	-0.735 (-4.991)	-0.739 (-4.451)	-0.727 (-5.621)	-0.714 (-5.423)
AIC [e]	405.76	387.85	387.67	388.58	375.292	378.50	374.78	373.79
$W^{sup}_{Type}\left(\hat{\theta}, \hat{d}, \hat{p}\right)$ [f]	NA	10.183	14.560	9.206	35.859	31.300	36.322	36.409
Linearity [g]	NA	3.256 (0.890) [h]	7.073 (0.992)	2.404 (0.758)	11.734 (0.970)	6.145 (0.091)	12.170 (0.822)	6.784 (0.004)
$Q(12)$ [i]	11.989 (0.447)	13.173 (0.356)	14.643 (0.262)	11.802 (0.462)	16.327 (0.177)	17.128 (0.145)	10.338 (0.586)	9.543 (0.656)

Given that the long- and short term interest rates are $I(1)$, stationarity of the interest rate differential implies the property of cointegration and, hence, the following threshold vector error correction representation

$$\Delta \mathbf{x}_t = \begin{cases} \mu^{(1)} + \boldsymbol{\alpha}^{(1)} z_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i^{(1)} \Delta \mathbf{x}_{t-i} + \boldsymbol{\eta}_t^{(1)} & if \quad \Delta z_{t-d} < \theta^{(1)} \\ \mu^{(2)} + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i^{(2)} \Delta \mathbf{x}_{t-i} + \boldsymbol{\eta}_t^{(2)} & if \quad \theta^{(1)} \le \Delta z_{t-d} \le \theta^{(2)} \\ \mu^{(3)} + \boldsymbol{\alpha}^{(3)} z_{t-1} + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i^{(3)} \Delta \mathbf{x}_{t-i} + \boldsymbol{\eta}_t^{(3)} & if \quad \Delta z_{t-d} > \theta^{(2)} \end{cases} \tag{5.29}$$

where $\mathbf{x}_t = \left[x_{St}, x_{Lt}\right]'$, letting x_{St} and x_{Lt} denote the short- and long-term interest rates, respectively, and z_{t-1} represents the lagged interest rate differential. Note that since the inner-regime in a Band-MTAR model is propelled by a unit root the *VECM* reduces to a stable $VAR(p-1)$ representation when $\theta^{(1)} \le \Delta z_{t-d} \le \theta^{(2)}$.

Table VIII
(Threshold)Vector Error Correction Model Estimates

[a] The Ljung-Box statistic testing whether the first 12 residual autocorrelations are jointly equal to 0. [b]A likelihood ratio test for testing the null hypothesis that all elements in the relevant coefficient matrix are jointly equal to zero. [c] Hosking's multivariate portmanteau test that the first 8 autocovariance matrices are jointly equal to 0. [d] The Akaike information criterion computed as $AIC = \ln\left|\Sigma_{\eta}\right| + \left(2/N^{*}\right) \times$ the number of freely estimated parameters. [e] likelihood ratio test when linearity has been imposed on the cointegrating residuals. Significance levels are in parentheses.

		Model for Δx_{St}		Model for Δx_{Lt}		
Regime	Explanatory Variables	Coefficient	t-statistic	Coefficient	t-statistic	
			Model Estimates of TVECM			
Lower	$z_{t-1} - \theta^{(l)}$	-0.358	-0.183	0.077	0.046	
Upper	$z_{t-1} - \theta^{(u)}$	-0.986	-0.376	-0.447	-0.197	$LRT(\Psi_3 = 0)^{b} = 38.18$ (0.00)
	Constant	0.061	0.041	-0.083	-0.064	
All	$\Delta x_{S,t-1}$	-0.713	-0.469	-0.104	-0.079	$LRT(\Psi_2 = 0) = 5.42$ (0.25)
	$\Delta x_{L,t-1}$	0.877	0.501	0.160	0.106	
		$Q(12)^{a} = 15.59$ (0.21)		$Q(12) = 13.97$ (0.30)		
			$H(32)^{c} = 58.88$ (0.00)			$AIC^{d} = 9.40$
			Model Estimates of a Linear VECM			
NA	z_{t-1}	-0.729	-0.522	-0.193	-0.163	$LRT(\alpha = 0) = 31.45$ (0.00)
	Constant	-0.138	-0.092	-0.124	-0.097	$LRT(\Psi_1 = 0) = 7.14$ (0.13)
		$Q(12) = 12.07$ (0.44)		$Q(12) = 12.96$ (0.37)		
			$H(28) = 50.13$ (0.24)			$AIC = 9.42$
						$\chi^{2\,e} = 14.17$ (0.00)

Table VIII presents parameter estimates of two error correction models. The first, labeled *TVECM*[82], is model (5.29) under the constraints $\mu^{(l)} = \mu^{(m)} = \mu^{(u)}$ and $\Psi_i^{(l)} = \Psi_i^{(m)} = \Psi_i^{(u)}$ for all i, while the second is a *linear VECM* model, entailing the imposition of the constraint $\alpha^{(l)} = \alpha^{(m)} = \alpha^{(u)}$ on *TVECM*[83].

[82] We estimated model (5.29) and found it to be unstable. This finding is unsurprising since the estimated thresholds of the Band-MTAR model partition the 89 usable cases into 18, 16 and 55 cases in the lower, middle and upper regimes respectively. Resultantly, we omit it from further discussion.

[83] Note that $\alpha^{(z)}$ has been omitted from (5.29) where we assume that the inner-band follows a unit root process implying $\alpha^{(z)} = 0$.

Utilizing the Akaike information criterion and assessing whether the maximized likelihood is significantly altered, by sequentially imposing the restriction $\Psi_i^{(j)} = \mathbf{0}$, for $i = m, m-1, \ldots, 0$, where m denotes an *a priori* choice of the maximum allowable lag-order, leads to the finding that *TVECM* has a single lagged value of the dependent variable while the linear VECM has none. *TVECM*, which, according to the χ^2 statistic, is significantly different from the linear VECM, appears to model the dynamics of the system more closely; the univariate Ljung-Box statistics, on aggregate, and Hosking's multivariate portmanteau statistic, similarly, suggest this finding. An important implication of these findings is that if thresholds are present in the cointegrating residuals error correction models which are unable to take this characteristic into account will be misspecified.

5.7. Conclusion

This chapter detailed a direct testing procedure of the unit root hypothesis by specifying momentum-TAR type asymmetric equilibrium-reversion under the alternative hypothesis. The findings of Cook (2003), illustrating that M-TAR type nonlinearity is more likely to be detected in a nonlinear process of unknown form, implies the need for a unit root test of the type that we propose. We presented critical values, in the trend of Dickey and Fuller (1979), by accounting for the possible presence of deterministic components, in the form of a constant and/or a trending set of attractors, and presented power and size tests when the process under consideration reverts to an equilibrium level and when it reverts back to an equilibrium band. While our test dominates conventional unit root tests in the form of the augmented Dickey-Fuller and Phillips-Perron tests, as well as, recent nonlinear tests represented by the two Enders-Granger and the Berben-van Dijk tests in terms of power it is prone to over-rejection. The sizable size distortion is however not unique since other tests, such as that of Berben and van Dijk (1998), display similar distortion.

By way of illustrating how the tests are applied we investigated the term structure of quarterly Eurocurrency interest rates for the U.K. Our findings suggest that the interest rate differential is not only asymmetric and stationary, as is documented by a number of recent studies, but also has the property of varying randomly when within an inner-band. The threshold vector error correction model implied by these findings delivers new insights into the dynamics that determine short-and long-term interest rates. Suggestions for future research include extending the approach to the case where the cointegrating vector is unknown;

possibly requiring optimization over the joint cointegrating vector and threshold parameter spaces.

6

Multivariate Threshold Cointegration: Forecasting the Spot Exchange Rate with the Term Structure of Forward Premiums

6.1. Introduction

Linear cointegration implicitly presupposes that reversion to long-run equilibrium is both continuous and symmetric, while economic conjectures abound as to why this need not *necessarily* be the case; the law of one price being a pertinent example. Consequentially, Balke and Fomby (1997) detail veridical equilibrium reversion processes that more accurately reflect the empirical findings insinuated by the studies of a plenitude of researchers from the many, diverse, sub-domains in economics. By allowing for the presence of thresholds in the disequilibrium process of a cointegrated system they further the general tendence of developing nonlinear generalizations of linear techniques; a deliberate break from Slutzky's (1927) linear paradigm.

Curiously, while a plenitude of researchers have developed statistical tests and procedures for the case of threshold cointegration in a bivariate system (see Balke and Fomby (1997), Enders and Granger (1998), Lo and Zivot (2001), Hansen and Seo (2002) and Seo (2003a, 2003b)) multivariate threshold cointegration has, as yet, been left untended. In this chapter we introduce the notion of multivariate threshold cointegration whereby each time-series of cointegrating residuals in a cointegrated system are governed by separate threshold

autoregressive models. Additionally, we present a simple and convenient formulation of the multivariate threshold VECM it implies.

Empirically, we assess the contribution of the model by applying it to the task of forecasting spot exchange rates using the term structure of forward premia. A theoretical foundation is given by the work of Clarida and Taylor (1997) who develop a framework for the prediction of future spot rates by extracting information from the term structure of forward exchange premiums. They do so by showing that while both the spot and forward rates are individually integrated processes an economically justifiable combination of these processes is stationary – implying the existence of the property of cointegration and the subsequent existence of a VECM that may be used for forecasting spot rates. While the empirical reassessment of their results presented in Chapter 2 was unable to validate their considerable forecasting results the theoretical framework on which their empirical study hinges is interesting and leads naturally to the idea of threshold cointegrated spot and forward rates.

§6.2 details and presents a general multivariate threshold cointegration framework. §6.3 ensuingly discusses a two-step estimation procedure whereby the threshold cointegrated system is estimated conditionally on parameter estimates of the threshold autoregressions which we hypothesis govern the sequence of disequilibria. §6.4 presents the theoretical framework of Clarida and Taylor (1997) which we extend to allow for the incorporation of threshold effects in the forward premia of a number of exchange rates. §6.5 delineates a comprehensive study of the stationarity and linearity properties of the forward premia. Once nonlinearity has been determined we estimate the multivariate threshold model these properties imply and utilize the model in producing out-of-sample forecasts which are juxtaposed against a series of alternative forecasting models. §6.6 concludes and provides a number of suggestions for future research.

6.2. A Multivariate Threshold Cointegration Framework

To see how multivariate threshold cointegration, of rank r, can arise consider a K-dimensional $VAR(p)$ model

$$\mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\Gamma}_1 \mathbf{x}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \mathbf{x}_{t-p} + \boldsymbol{\eta}_t \tag{6.1}$$

where $\mathbf{x}_t = (x_{1,t}, x_{2,t}, \ldots, x_{K,t})$, for $t = 1,2,\ldots,N$, with $\boldsymbol{\eta}_t$ an innovation process i.e. $E[\boldsymbol{\eta}_t] = \mathbf{0}$, $E[\boldsymbol{\eta}_t \boldsymbol{\eta}_t'] = \boldsymbol{\Sigma}_\eta$, and $E[\boldsymbol{\eta}_t \boldsymbol{\eta}_s'] = \mathbf{0}$ for $s \neq t$. Suppose that the process is nonstationary with $|\mathbf{I}_2 - \boldsymbol{\Gamma}_1 z - \cdots - \boldsymbol{\Gamma}_p z^p| = (1-\lambda_1 z)\cdots(1-\lambda_n z) = 0$ for $z = 1$, where λ_i are the reciprocals of the roots of the determinantal polynomial. Logically, at least one of them must equal unity; while all other roots are assumed to lie outside the unit circle, i.e. all $\lambda_i \neq 1$ are inside the complex unit circle. Then since $|\mathbf{I}_2 - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_p| = 0$ the matrix $\boldsymbol{\Pi} = \mathbf{I}_2 - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_p$ is singular and if $rank(\boldsymbol{\Pi}) = r$, it may be decomposed as $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}$, where $\boldsymbol{\alpha}$ is $(K \times r)$ and $\boldsymbol{\beta}$ is $(r \times K)$. The matrix $\boldsymbol{\alpha}$ is typically referred to as the loading matrix while $\boldsymbol{\beta}$ is the matrix of cointegrating vectors. If $r = 0$, $\Delta\mathbf{x}_t$ will have a stable $VAR(p-1)$ representation and, if $r = K$, $\mathbf{x}_t$ is a stationary $VAR(p)$ process.

If $0 < rank(\boldsymbol{\Pi}) = r < K$, model (6.1) has the following error correction representation

$$\Delta\mathbf{x}_t = \boldsymbol{\mu} - \boldsymbol{\alpha}\boldsymbol{\beta}\mathbf{x}_{t-1} + \boldsymbol{\Psi}_1 \Delta\mathbf{x}_{t-1} + \cdots + \boldsymbol{\Psi}_{p-1}\Delta\mathbf{x}_{t-p+1} + \boldsymbol{\eta}_t \tag{6.2}$$

where $\boldsymbol{\Psi}_i = -(\boldsymbol{\Gamma}_{i+1} + \cdots + \boldsymbol{\Gamma}_p)$ for $i = 1,\ldots,p-1$[84]. Let, allowing for the simplifying assumption that each equilibrium error process is independent of all others, the *r*-vector $\mathbf{z}_t = [z_{1,t}, z_{2,t}, \ldots, z_{r,t}]$, where $z_{i,t} = \beta_i x_t$ with β_i denoting the *ith* row vector of $\boldsymbol{\beta}$, define *r*-discrepancies from long-run equilibria and, additionally, allow each $z_{i,t}$ to have a *k*-regime TAR representation[85], i.e.

$$z_{i,t} = \sum_{j=1}^{k} \left[\phi_{i,0}^{(j)} + \phi_{i,1}^{(j)} z_{it-1} + \cdots + \phi_{i,p^{(j)}}^{(j)} z_{i,t-p} + \varepsilon_{i,t}^{(j)}\right] I\left(\theta_i^{(j-1)} < v_{i,t-d} \leq \theta_i^{(j)}\right) \tag{6.3}$$

where the positive integers *j*, *d*, and *p* denote regime index, threshold delay and autoregressive lag-order, respectively; $\varepsilon_{i,t}^{(j)}$ is assumed to be a martingale difference sequence with respect to the past history of z_t; $\phi_{i,l}^{(j)}$ denotes the autoregressive coefficient of the *lth* lag in regime *j*; $I(A)$ is an indicator function such that it equals unity when event *A* occurs and is

[84] See Hylleberg and Mizon (1989) for other representations of cointegrated processes.
[85] In the context of the framework that we develop the equilibrium error processes may be assigned any functional form; including an allowance for the interplay between two or more disequilibrium processes.

zero otherwise; $v_{i,t-d}$ is the delay variable which is evaluated relative to the threshold vector $\boldsymbol{\theta}_i = \left[\theta_i^{(0)}, \theta_i^{(1)}, \ldots, \theta_i^{(k)}\right]$, where $\theta_i^{(k)} = -\theta_i^{(0)} = \infty$. If $v_{i,t-d} \equiv z_{i,t-d}$ then (6.3) is typically referred to as a self-exciting TAR (SETAR) model.

Conditional on the existence of k non-trivial thresholds, in (6.3), a general multivariate threshold error correction representation may be formulated as

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}^{(1)} \mathbf{I}^{(1)} \mathbf{z}_{t-1}^{(1)} + \boldsymbol{\alpha}^{(2)} \mathbf{I}^{(2)} \mathbf{z}_{t-1}^{(2)} + \cdots + \boldsymbol{\alpha}^{(k)} \mathbf{I}^{(k)} \mathbf{z}_{t-1}^{(k)} + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\eta}_t \tag{6.4}$$

or, more compactly, as

$$\Delta \mathbf{x}_t = \boldsymbol{\mu} + \sum_{j=1}^{k} \boldsymbol{\alpha}^{(j)} \mathbf{I}^{(j)} \mathbf{z}_{t-1}^{(j)} + \sum_{i=1}^{p-1} \boldsymbol{\Psi}_i \Delta \mathbf{x}_{t-i} + \boldsymbol{\eta}_t \tag{6.5}$$

where $\boldsymbol{\alpha}^{(j)}$ is the $(K \times r)$ loading matrix associated with equilibrium errors $\mathbf{z}_t^{(j)}$ when $\theta_i^{(j-1)} < v_{i,t-d} \leq \theta_i^{(j)}$, for all $i = 1, \ldots, r$; $\mathbf{I}^{(j)}$ is a $(r \times rk)$ block diagonal indicator matrix, the *jth* element of the main diagonal being equal to unity when $\theta_i^{(j-1)} < v_{i,t-d} \leq \theta_i^{(j)}$ and is zero otherwise with all off-diagonal elements set equal to zero; $\mathbf{z}_{t-1}^{(j)}$ is a lagged-value of the disequilibrium term being explicitly dependent on the functional form present in the *jth* regime of the *ith* cointegrating residual[86].

To further clarify the formulation of model (6.5) we briefly digress to discuss an example. Suppose a trivariate cointegrated system, propelled by a single common trend, with each time-series of cointegrating residuals being governed by a double-threshold Band-TAR model. A Band-TAR model describes the following time-series process

$$z_t = \begin{cases} \sum_{i=1}^{p} \phi_i^{(1)} \left(z_{t-i} - \theta^{(1)}\right) + \varepsilon_t & \text{if} \quad z_{t-d} < \theta^{(1)} \\ \mu + \sum_{i=1}^{p} \phi_i^{(2)} z_{t-i} + \varepsilon_t & \text{if} \quad \theta^{(1)} \leq z_{t-d} \leq \theta^{(2)} \\ \sum_{i=1}^{p} \phi_i^{(3)} \left(z_{t-i} - \theta^{(2)}\right) + \varepsilon_t & \text{if} \quad z_{t-d} > \theta^{(2)} \end{cases} \tag{6.6}$$

[86] Note that the Threshold VECM, or TVECM, typically discussed in the literature relates only to cointegration in a bivariate system and, hence, defines a model that is very different from model (6.5). A bivariate TVECM (see Lo and Zivot (2001) for a detailed discussion) has the advantage of allowing the lagged changes in a VECM to be determined by the threshold variable of the TAR model governing the series of cointegrating residuals; while the multivariate setting does not possess this capacity.

or, alternatively, using the indicator function $I^{(j)}(A)$, which is equal to unity when event A occurs and is zero otherwise, (6.6) may be written as

$$z_t = I^{(1)}\left(z_{t-d} < \theta^{(1)}\right)\left[\sum_{i=1}^{p}\phi_i^{(1)}\left(z_{t-i} - \theta^{(1)}\right)\right] + I^{(2)}\left(\theta^{(1)} \le z_{t-d} \le \theta^{(2)}\right)\left[\mu + \sum_{i=1}^{p}\phi_i^{(2)}z_{t-i}\right] + I^{(3)}\left(z_{t-d} > \theta^{(2)}\right)\left[\sum_{i=1}^{p}\phi_i^{(3)}\left(z_{t-i} - \theta^{(2)}\right)\right] + \varepsilon_t \tag{6.7}$$

Hence, defining

$$\mathbf{I} = \left[I^{(1)}, I^{(2)}, I^{(3)}\right] \tag{6.8}$$

as a $(1 \times k)$ indicator vector, where $I^{(j)} = I^{(j)}(A)$, we can reformulate (6.7) compactly as

$$z_t = \mathbf{I}\mathbf{f}\left(z_{t-}\right) \tag{6.9}$$

where

$$\mathbf{f}\left(z_{t-}\right) = \begin{bmatrix} \sum_{i=1}^{p}\phi_i^{(1)}\left(z_{t-i} - \theta^{(1)}\right) + \varepsilon_t \\ \mu + \sum_{i=1}^{p}\phi_i^{(2)}z_{t-i} + \varepsilon_t \\ \sum_{i=1}^{p}\phi_i^{(3)}\left(z_{t-i} - \theta^{(2)}\right) + \varepsilon_t \end{bmatrix} \tag{6.10}$$

is a vector allowing for the incorporation of the different functional forms present in each of the *k*-regimes of the model. Denoting the *ith* disequilibrium process $z_{i,t}$, for $i = 1,2$, given the presence of a single common trend (see Stock and Watson (1988)), we can simply extend the formulation in (6.8) to allow for multiple cointegrating residual series. Let $\mathbf{I}_i$ denote vector (6.8) and $\mathbf{f}_i\left(z_{t-}\right)$ denote vector (6.10) for the *ith* disequilibrium term, then

$$\mathbf{z}_t = \begin{bmatrix} z_{1,t} \\ z_{2,t} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1\mathbf{f}_1\left(z_{t-}\right) \\ \mathbf{I}_2\mathbf{f}_2\left(z_{t-}\right) \end{bmatrix} = \begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix}\begin{bmatrix} \mathbf{f}_1\left(z_{t-}\right) \\ \mathbf{f}_2\left(z_{t-}\right) \end{bmatrix}$$

and since $\begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} = \begin{bmatrix} I_1^{(1)} & I_1^{(2)} & I_1^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & I_2^{(1)} & I_2^{(2)} & I_2^{(3)} \end{bmatrix}$ it is useful to decompose the matrix according to which regime $z_{i,t}$ is in i.e.

$$\begin{bmatrix} \mathbf{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_2 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(z_{t-}) \\ \mathbf{f}_2(z_{t-}) \end{bmatrix} = \begin{bmatrix} I_1^{(1)} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_2^{(1)} & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(z_{t-}) \\ \mathbf{f}_2(z_{t-}) \end{bmatrix} + \begin{bmatrix} 0 & I_1^{(2)} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_2^{(2)} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(z_{t-}) \\ \mathbf{f}_2(z_{t-}) \end{bmatrix} + \begin{bmatrix} 0 & 0 & I_1^{(3)} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & I_2^{(3)} \end{bmatrix} \begin{bmatrix} \mathbf{f}_1(z_{t-}) \\ \mathbf{f}_2(z_{t-}) \end{bmatrix} \tag{6.11}$$

Hence, denoting the matrix corresponding with regime j in 6.11 by $\mathbf{I}^{(j)}$ and $\mathbf{z}_t = \begin{bmatrix} \mathbf{f}_1(z_{t-}) \\ \mathbf{f}_2(z_{t-}) \end{bmatrix}$ we have that (6.11) may be conveniently formulated as $\sum_{j=1}^{k} \mathbf{I}^{(j)} \mathbf{z}_t$ and the error correction term $\boldsymbol{\beta} \mathbf{x}_{t-1}$, in (6.2), may be substituted with $\sum_{j=1}^{k} \mathbf{I}^{(j)} \mathbf{z}_{t-1}$ leading to the formulation in model (6.5).

Stationarity of (6.5) is ensured if the outer-regimes of $\mathbf{z}_t$ are stationary; Tjøstheim (1990) has shown that a sufficient condition for stationarity of (6.3), and hence (6.5), is that the roots of the AR processes in the outer-regimes are less than $|1|$ [87]. Unit root testing in the presence of threshold-nonlinearity requires a model under the alternative that takes account of this characteristic; Pippenger and Goering (1993) were the first to show the severely detrimental effect threshold boundaries have on conventional unit root tests. Consequently, Enders and Granger (1998), Berben and van Dijk (1999) and Kapetanios and Shin (2002) have developed tests specifying single and double-threshold models under the alternative. In Chapter 5 we developed a unit root test, with superior power to these alternatives.

In this section we developed a general multivariate threshold error correction model with the capacity of allowing differing threshold processes to define each cointegrating residual series; including the special case of linear cointegration, when trivial threshold effects are present in each series, as well as, no cointegration when, for example, each disequilibrium term is an element of the inner-regime of a Band-TAR process admitting a unit root process

[87] See §2.3.2 for a more detailed discussion of the conditions required to ensure stationarity of a TAR process.

there within. Additionally, the formulation is conducive to standard estimation procedures; the specifics of which are detailed in the next section.

6.3. Multivariate Threshold VECM Estimation

6.3.1 TAR Model Parameter Estimation

The general multivariate threshold VECM developed in the previous section convolutes around two estimation problems. The first relates to specifying and estimating TAR models for each series of cointegrating residuals while the second entails estimation of the autoregressive parameters of the VECM, conditional on the TAR model estimates. Disentangling these estimations seems the most tractable approach; direct estimation being in an embryonic stage of development relates primarily to bivariate cointegrated systems. Resultantly, the estimation procedures which we detail in this section relate firstly to TAR model estimation and then the conditionally estimated parameters of model (6.5).

The range of possible specifications of model (6.3) is richly diverse. Consider the following two "first-difference" reparameterizations of (6.3)

$$\begin{aligned}\Delta z_t = &\left[\phi_0^{(1)} + \phi_1^{(1)} z_{t-1}\right] I_t\left(z_{t-1} < \theta^{(1)}\right) + \left[\phi_0^{(2)} + \phi_1^{(2)} z_{t-1}\right] I_t\left(\theta^{(1)} \le z_{t-1} < \theta^{(2)}\right) + \\ &\left[\phi_0^{(3)} + \phi_1^{(3)} z_{t-1}\right] I_t\left(z_{t-1} \ge \theta^{(2)}\right) + \varepsilon_t\end{aligned} \tag{6.12}$$

$$\begin{aligned}\Delta z_t = &\left[\phi_1^{(1)} (z_{t-1} - \theta^{(1)})\right] I_t\left(z_{t-1} < \theta^{(1)}\right) + \left[\phi_0^{(2)} + \phi_1^{(2)} z_{t-1}\right] I_t\left(\theta^{(1)} \le z_{t-1} < \theta^{(2)}\right) \\ &\left[\phi_1^{(3)} (z_{t-1} - \theta^{(2)})\right] I_t\left(z_{t-1} \ge \theta^{(2)}\right) + \varepsilon_t\end{aligned} \tag{6.13}$$

where, for convenience, it is assumed that $\varepsilon_t \sim IIN\left(0, \sigma_\varepsilon^2\right)$ while Δ denotes the first-difference operator i.e. $\Delta z_t = z_t - z_{t-1}$. Notationally, the remainder may be inferred from (6.3).

Equations (6.12) and (6.13) are generalizations of models that are typically referred to as Equilibrium-TAR and Band-TAR models; see Balke and Fomby (1997). Here we refrain from instituting the requirement that the thresholds are symmetrically dispersed around the unconditional mean of the cointegrating residual process, nor do we require that the AR processes in the outer-regimes are equivalent. Enders and Granger (1998) introduced the idea of using momentum as a delay variable by substituting z_{t-d} in a SETAR model with Δz_{t-1};

terming the resultant model a Momentum-TAR, or M-TAR, model. Their findings, when specifying an M-TAR model under the alternative of a unit root test, indicate the potential contribution that the model could have when allowing discrepancies from equilibrium in a threshold cointegrated system to be modeled by an M-TAR model. Subjecting (6.4) to the requirement that the piecewise linear autoregressive function be continuous everywhere, results in an important TAR sub-class, the continuous TAR or CTAR model, introduced into the literature by Chan and Tsay (1998)[88].

Balke and Fomby (1997) further restrict specifications (6.12) and (6.13) by allowing the inner-regime to be driven by a unit root process. The consequences of doing so for (6.5) are interesting since then the process will follow a $VAR(p)$ model when the cointegrating residuals are within the corridor formed by the threshold vector of (6.4). When in the outer-regimes, however, $\mathbf{\Pi} = \boldsymbol{\alpha}^{(j)}\boldsymbol{\beta}$ will be of reduced-rank implying that an equilibrium reverting mechanism governs the fluctuating cointegrating residuals. Hence, (6.5) switches between a $VAR(p)$ and a VECM specification depending on the magnitude of the delay variable relative to the locations of the thresholds in (6.3).

Generally, TAR models may be categorized as belonging to one of two sub-classes: those that have regimes where the AR process is a functional of a threshold parameter and those that do not. Fitting SETAR models with regimes that are independent of the values of the threshold vector is typically done by sequential conditional least squares. In this case the continuous threshold parameter space $\Theta \subseteq \Re^{k-1}$ is restricted to the set of points

$$\mathbf{\Theta} = \left\{ \boldsymbol{\theta} \mid z_{\left(\lfloor N^* \pi_0 \rfloor\right)} \leq \theta^{(j)} \leq z_{\left(\lfloor (1-\pi_1) N^* \rfloor\right)} \right\} \tag{6.14}$$

where $z_{(i)}$ denotes the *ith* order statistic of the delay variable; (π_0, π_1) are trimming parameters which ensure that an adequate number of observations are present in the outer-regimes to ensure the accurate estimation of the parameters of that regime; $N^* = N - \max(d, L)$ is the effective sample size where L is an *a priori* choice of the maximum allowable dimension of the lag space; and $\lfloor \cdot \rfloor$ denotes a floor function here defined

[88]Instituting the restrictions $p = 1$; $\phi_0^{(2)} = 0$; and $\phi_1^{(2)} = 1$ results in three-regime SETAR models that allow for unit root behavior in the inner-regime and autoregressive behavior in the outer-regimes. It is easily shown that restricting (6.13), in this manner, will result in a model that meets the "continuous everywhere" requirement implying that it belongs to the CTAR sub-class.

as the integer operator. Parameter estimation is done by computing the residual variance conditional on the *gth* candidate threshold vector as

$$\tilde{\sigma}_g^2 = \frac{1}{N^*}\sum_{j=1}^{k}\left(\Delta \mathbf{z}_g^{(j)} - \mathbf{F0}_g^{(j)}\tilde{\mathbf{\Phi}}_g^{(j)}\right)'\left(\Delta \mathbf{z}_g^{(j)} - \mathbf{F0}_g^{(j)}\tilde{\mathbf{\Phi}}_g^{(j)}\right) \tag{6.15}$$

where $\Delta \mathbf{z}_g^{(j)}$ denotes the dependent variable in regime *j*, in (6.12) for example; and $\mathbf{F0}_g^{(j)}$ is the $\left(N^{*(j)} \times L\right)$ matrix of regressor values, where $N^{*(j)}$ is the effective sample size in regime *j*[89]; and the vector of autoregressive coefficient estimates are defined as $\tilde{\mathbf{\Phi}}_g^{(j)} = \left[\tilde{\phi}_1^{(j)}, \tilde{\phi}_2^{(j)}, \ldots, \tilde{\phi}_L^{(j)}\right]$ where $\tilde{\phi}_i^{(j)}$ is the coefficient estimate of the *ith* lag in regime *j*. Letting *G* denote the total number of candidate threshold vectors the estimated threshold vector is given by $\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_{\hat{g}}$ where $\hat{g} = \underset{1 \le g \le G}{\arg\min}\, \tilde{\sigma}_g^2$, while $\hat{\mathbf{\Phi}}^{(j)} = \hat{\mathbf{\Phi}}_{\hat{g}}^{(j)} = \left(\mathbf{F0}_{\hat{g}}^{(j)'}\mathbf{F0}_{\hat{g}}^{(j)}\right)^{-1}\left(\mathbf{F0}_{\hat{g}}^{(j)'}\Delta \mathbf{z}\right)$.

Allowing $p^{(j)} < L$ requires the use of information criteria, such as the Akaike Information Criterion (*AIC*) measure of Tong (1983), to determine the appropriate lag-order in each regime[90]. Similarly the normalized AIC (*N AIC*), as suggested by Tong and Lim (1980), is typically utilized to determine the value of the unknown threshold delay. Chan (1993) has shown that $\hat{\boldsymbol{\theta}}$ is asymptotically independent of $\hat{\mathbf{\Phi}}^{(j)}$, for all *j*, and that the latter estimates are normally distributed, as is the case when the threshold vector is known. This is a consequence of his illustration that, for arbitrary *L*, the conditional LS estimates of the parameters of a $TAR(p;d;k)$ model are strongly consistent.

If the regressor matrix may be written as the first-degree polynomial matrix $\mathbf{F}_g^{(j)} = \mathbf{F0}_g^{(j)} + \mathbf{U}_g^{(j)}$, where $\mathbf{U}_g^{(j)} = \left[c, c, \ldots, c\right]'\left[\theta_g^{(i)}, \theta_g^{(i)}, \ldots, \theta_g^{(i)}\right]$ for *c*, being a scalar, and $\theta_g^{(i)}$, being the threshold on which the AR process in regime j is explicitly dependent, then the continuous threshold parameter space will have to be searched to obtain the optimal threshold vector location. The task is relatively simple when the threshold vector is a scalar and may be computed efficiently by utilizing the fitting procedure proposed by Coakley *et al.* (2003). The fitting approach that they advocate involves subjecting an interpolated residual sum of squares function to the standard optimality condition. The support points required by the interpolating algorithm are computed via Givens transformations of QR factorizations of the (augmented)

[89] Frequently augmented with a unit vector when including a constant term. For reasons of notational convenience we abstain from including it here.

[90] Alternatively, the more parsimonious Schwartz Bayesian criterion (SBC) or the Hannan-Quinn criterion (HQC) may be used (see Kapetanios (1999)).

regressor matrix $\left[\mathbf{F0}_g^{(j)} \mid \Delta\mathbf{z}_g^{(j)}\right]$. When generalizing the fitting approach to allow for multiple thresholds in a TAR process, as we did in Chapter 4, the LS fitting problem becomes decidedly less trivial; involving the need for a multidimensional interpolation-based optimization routine defined over a set of pedantically defined non-overlapping threshold hypercubes.

6.3.2 Conditional Multivariate VECM Parameter Estimation

Given the r TAR model estimates, discussed in the previous section, parameter estimation of the multivariate threshold VECM is fairly straightforward. The estimation procedure results in sets of threshold vector estimates, here, denoted by $\hat{\boldsymbol{\theta}}_i = \left[\hat{\theta}_i^{(0)}, \hat{\theta}_i^{(1)}, \ldots, \hat{\theta}_i^{(k)}\right]$, for each of the r cointegrating residual series. Conditional on these estimates let $\mathbf{I}_M^{(j)}$ denote a $N^* \times r$ matrix of indicator vectors, for $j = 1,2,\ldots,k$, then denoting the *ith* element of the *tth* row vector by $\mathbf{I}_M^{(j)}(t,i)$ we have

$$\mathbf{I}_M^{(j)}(t,i) = \begin{cases} 1 & \text{if} \quad \hat{\theta}_i^{(j-1)} < v_{i,t-d} \le \hat{\theta}_i^{(j)} \\ 0 & \text{otherwise} \end{cases} \tag{6.16}$$

Then $\mathbf{I}_M^{(j)} \circ \mathbf{z}'$, where $\mathbf{z}$ is a $r \times N^*$ matrix in which the *ith* element of each column vector is defined as in (6.9) for $t = 1,2,\ldots,N^*$ and $\circ$ denotes a Hadamard product[91].

Subsequently, if we denote by $\mathbf{Z}$ the horizontally concatenated matrix of regressor values i.e. $\left[1 \mid \mathbf{I}_M^{(1)} \circ \mathbf{z}_{i,t-1}^{(1)\prime} \mid \mathbf{I}_M^{(2)} \circ \mathbf{z}_{i,t-1}^{(2)\prime} \mid \mathbf{I}_M^{(3)} \circ \mathbf{z}_{i,t-1}^{(3)\prime} \mid \Delta\mathbf{x}_{t-1} \mid \cdots \mid \Delta\mathbf{x}_{t-p+1}\right]$ OLS estimates of the coefficients, $\mathbf{B}$, are obtained as $\hat{\mathbf{B}} = \mathbf{Y}\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\mathbf{Y} = \left[\Delta\mathbf{x}_1, \Delta\mathbf{x}_2, \ldots, \Delta\mathbf{x}_{N^*}\right]$ and an estimate of the residual covariance matrix is given by $\Sigma_\eta = \mathbf{Y}\left(I_{N^*} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}\right)\mathbf{Y}'$ with I_{N^*} an identity matrix of order N^*. Finally, Lütkepohl (1993) provides a detailed discussion on VECM order selection and diagnostic tests of model adequacy applicable to the case at hand. In the section that follows we introduce the empirical framework of Clarida and Taylor (1997) and detail the applicability of the underlying notions of model (6.5) to that context.

---

[91] If A and B are matrices of the same order, say $m \times n$, with elements a_{ij} and b_{ij}, respectively, then the $m \times n$ matrix $A \circ B$ defines a Hadamard product when its *ijth* element is $a_{ij}b_{ij}$.

6.4. Forecasting Spot Exchange Rates in a Multivariate Threshold Cointegration Framework

6.4.1 The Clarida-Taylor Empirical Framework

Clarida and Taylor (1997) develop an empirical framework, drawing upon a similar framework employed by Hall *et al.* (1992) in a study of the term structure of treasury bill yields, in which a 5-variable system of spot and forward exchange rates are hypothesized to be propelled by a single common trend. We investigate, discretely sampled, weekly data on spot and 4-, 13-, 26-, 52-week forward dollar exchange rates for Germany, Japan, and the United Kingdom. The data span the period 1977:1 through 1993:52; model estimation is carried out on the sub-sample 1977:1 through 1990:26 leaving all complimentary data for the dynamic out-of-sample forecasting exercises.

Below we briefly delineate the Clarida and Taylor (1997) framework after which we introduce an alternative formulation of the short-run dynamics of the process.

Let

$$\mathbf{y}_t = \left[s_t, f_{h(1),t}, f_{h(2),t}, \ldots, f_{h(5),t}\right]' \tag{6.17}$$

where s_t and $f_{h(\tau),t}$ denote the logarithm of the spot and forward exchange rates at horizon $h(\tau)$, respectively. Assuming the empirically pervasive finding of a unit in the spot rate it is allowed to evolve according to

$$s_t = w_t + v_t \tag{6.18}$$

where $v_t \sim I(0)$ and $w_t = \gamma + w_{t-1} + e_t$. Letting $E(s_{t+h(\tau)} \mid \Omega_t)$ denote the mathematical expectation of the spot rate at horizon $h(\tau)$, conditional on the information that is available at time t, Ω_t, then any deviation from the risk-neutral efficient markets hypothesis maybe defined as

$$\delta_{h(\tau),t} \equiv f_{h(\tau),t} - E(s_{t+h(\tau)} \mid \Omega_t) \tag{6.19}$$

Combining (6.18) and (6.19) results in the following expression of the forward rate at horizon $h(\tau)$

$$f_{h(\tau),t} = h(\tau)\gamma + w_t + E(v_{t+h(\tau)} \mid \Omega_t) + \delta_{h(\tau),t} \tag{6.20}$$

and subsequently to

$$f_{h(\tau),t} - s_t = h(\tau)\gamma + E(v_{t+h(\tau)} - v_t \mid \Omega_t) + \delta_{h(\tau),t} \tag{6.21}$$

which is the forward premium at horizon $h(\tau)$. Clearly, if the assumption in (6.19) holds each of the *j* forward premia are stationary implying the existence of a single common trend.

The authors subsequently assume that the equilibrium reversion process implicit in (6.19) is both symmetric and continuous while this need not necessarily be the case. It is conceivable that economic agents are unresponsive to random fluctuations in $\delta_{h(\tau),t}$, possibly for a prolonged period, yet have an economic incentive to act when these exceed some threshold level. Hence, the forward premia may be locally non-stationary while maintaining the global property of stationarity; in the sense of Balke and Fomby (1997).

Assuming that deviations from RNEMH are stationary and equilibrium reversion is guided by the following mechanism

$$\delta_{h(\tau),t} = \begin{cases} \alpha^{(3)} + \beta^{(3)}\left(\delta_{h(\tau),t-1} - \theta^{(2)}\right) + \varepsilon_t & \text{if} \quad \delta_{h(\tau),t-1} > \theta^{(2)} \\ \delta_{h(\tau),t-1} + \varepsilon_t & \text{if} \quad \theta^{(1)} < \delta_{h(\tau),t-1} \le \theta^{(2)} \\ \alpha^{(1)} + \beta^{(1)}\left(\delta_{h(\tau),t-1} - \theta^{(1)}\right) + \varepsilon_t & \text{if} \quad \delta_{h(\tau),t-1} \le \theta^{(1)} \end{cases} \tag{6.22}$$

which is a Band-TAR model the following model is implied for the forward premia

$$f_{h(\tau),t} - s_t = h(\tau)\gamma + \begin{cases} \alpha^{(3)} + \beta^{(3)}\left(\delta_{h(\tau),t-1} - \theta^{(2)}\right) + E(v_{t+h(\tau)} - v_t \mid \Omega_t) + \varepsilon_t & \text{if} \quad \delta_{h(\tau),t-1} > \theta^{(2)} \\ \delta_{h(\tau),t-1} + E(v_{t+h(\tau)} - v_t \mid \Omega_t) + \varepsilon_t & \text{if} \quad \theta^{(1)} < \delta_{h(\tau),t-1} \le \theta^{(2)} \\ \alpha^{(1)} + \beta^{(1)}\left(\delta_{h(\tau),t-1} - \theta^{(1)}\right) + E(v_{t+h(\tau)} - v_t \mid \Omega_t) + \varepsilon_t & \text{if} \quad \delta_{h(\tau),t-1} \le \theta^{(1)} \end{cases} \tag{6.23}$$

if we additionally assume that $\varepsilon_t \sim N\left(0,\sigma_\varepsilon^2\right)$ then if $\eta_t \equiv \varepsilon_t + E(v_{t+h(\tau)} - v_t \mid \Omega_t)$, $\eta_t \sim N(0,\sigma_\eta^2)$ where $0 < \sigma_\eta^2 < \infty$ (6.23) we may simplify model (6.22) as

$$f_{h(\tau),t} - s_t = h(\tau)\gamma + \begin{cases} \alpha^{(3)} + \beta^{(3)}\left(\delta_{h(\tau),t-1} - \theta^{(2)}\right) + \eta_t & \text{if} \quad \delta_{h(\tau),t-1} > \theta^{(2)} \\ \delta_{h(j),t-1} + \eta_t & \text{if} \quad \theta^{(1)} < \delta_{h(\tau),t-1} \le \theta^{(2)} \\ \alpha^{(1)} + \beta^{(1)}\left(\delta_{h(\tau),t-1} - \theta^{(1)}\right) + \eta_t & \text{if} \quad \delta_{h(\tau),t-1} \le \theta^{(1)} \end{cases} \tag{6.24}$$

Specifications (6.22) through (6.24) are stationary processes admitting a unit root when the process is in the inner-corridor and autoregressive equilibrium reversion when in an outer-regime. Interestingly, the thresholds of this model act as attractors hence we, implicitly, assume that the equilibrating process is towards the inner-band instead of towards a particular equilibrium level. Note that specification (6.24) allows for asymmetry in both the dispersion of the thresholds around the unconditional mean of the process, as well as, differing speeds of equilibrium reversion; the model is therefore more general than that originally proposed by Balke and Fomby (1997).

6.4.2 Nonlinearity and Stationarity of Forward Premiums

Postulating that each of the τ forward premiums possesses the properties of stationarity and nonlinearity requires an investigation into these properties. Since the testing strategy for linearity and a nonlinear unit root conventionally require the use of the supremum-Wald statistic it is convenient to state each hypothesis separately and then detail the testing procedure. A general formulation of the linearity hypothesis may be stated as

$$H_0 : \mathbf{\Phi}^{(j_1)} = \mathbf{\Phi}^{(j_2)} \text{ vs. } H_1 : \mathbf{\Phi}^{(j_1)} \neq \mathbf{\Phi}^{(j_2)} \tag{6.25}$$

where j_1 and j_2 represent the indices of all possible subsets of contiguous regimes with $j_1, j_2 \in \{1,2,\ldots,k\}$, while the unit root hypothesis may be stated as

$$H_0 : \gamma_1^{(j_1)} = \gamma_1^{(j_2)} = 0 \text{ vs. } H_1 : \gamma_1^{(j_1)} < 0 \quad \textit{and} \quad \gamma_1^{(j_2)} < 0 \tag{6.26}$$

where, *here*, j_1 and j_2 represent the indices of the outer-regimes with $j_1, j_2 \in \{1,2,\ldots,k\}$ and $\gamma_1^{(j_*)} = \phi_1^{(j_*)} - 1$, given the general TAR formulation in (6.3).

If the threshold locations are known, testing hypotheses (6.25) and (6.26) is achieved by means of computing the Wald statistic

$$W = \frac{N^*}{2}\left(\frac{\tilde{\sigma}^2 - \hat{\sigma}^2}{\hat{\sigma}^2}\right) \tag{6.27}$$

where $\tilde{\sigma}^2$ is the estimated residual variance under the null and $\hat{\sigma}^2$ is computed as is done in (6.15), conditional on the known threshold vector. If $\boldsymbol{\theta}$ is unknown it is appropriate to utilize the supremum-Wald test statistic, given estimates $\hat{d}$ and $\hat{p}$, which may be formulated as

$$W^{\sup}(\hat{d}, \hat{p}) = \sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} W(\boldsymbol{\theta}, \hat{d}, \hat{p}) \tag{6.28}$$

where

$$W(\boldsymbol{\theta}, \hat{d}, \hat{p}) = \frac{N^*}{2}\left(\frac{\tilde{\sigma}_n^2(\hat{p}) - \hat{\sigma}^2(\boldsymbol{\theta}, \hat{d}, \hat{p})}{\hat{\sigma}^2(\boldsymbol{\theta}, \hat{d}, \hat{p})}\right) \tag{6.29}$$

and $\hat{\sigma}^2(\boldsymbol{\theta}, \hat{d}, \hat{p})$ is the residual variance given some candidate threshold vector; defined such that the elements are equal to an order statistic of delay variable for models like (6.12). As alluded to previously, estimating the F-statistic for model (6.13) is more complex as it requires that each non-overlapping threshold hypercube[92] be searched by subjecting the RSS function, which is rational and continuous there within[93], to the standard optimality condition. Since the threshold parameter is unidentified under the null the asymptotic distribution of the sup-Wald test statistic is non-standard, consequently when testing (6.25), significance levels are computed using 200 replications of Hansen's (1996, 1999) fixed-regressor bootstrap. Significance of the unit root hypothesis is determined by comparing the value computed in (6.28) with the critical values computed by various authors given different stationary (SE)TAR model specifications under the alternative hypothesis.

[92] This construct is extensively detailed in §4.5.
[93] The proof of this statement can be found in the appendix of Coakley *et al.* (2003).

Table I
Linearity Tests of the Dollar-Sterling, Dollar-Mark and Dollar-Yen Forward Premiums.

*indicates rejection of the linearity hypothesis. p-values are computed using Hansen's (1996, 1999) fixed-regressor bootstrap with 200 replications.

Horizon	Single-Threshold non-linearity Supremum-Wald	Significance level	Double-Threshold non-linearity Supremum-Wald	Significance Level
Linearity Tests of the Forward Premia: Dollar-Sterling				
1-month	6.904	1.000*	80.931	0.000
3-month	1.355	0.635*	3.324	0.140*
6-month	0.835	0.335*	9.006	0.000
12-month	0.910	0.335*	6.190	0.010
Linearity Tests of the Forward Premia: Dollar-Mark				
1-month	15.712	1.000*	0.000	0.000
3-month	3.469	0.880*	59.000	0.000
6-month	3.373	0.815*	3.648	0.000
12-month	2.930	0.765*	8.258	0.020
Linearity Tests of the Forward Premia: Dollar-Yen				
1-month	3.293	0.850*	7.824	0.000
3-month	0.832	0.095*	1.495	0.000
6-month	0.955	0.085*	3.737	0.000
12-month	0.402	0.000	1.437	0.000

Tables I and II present the results of the linearity and unit root testing procedures, respectively. The results in Table I present a very clear image of the type of nonlinearity present in each of the forward premiums; being unable to reject the linearity hypothesis when testing for double-threshold nonlinearity while being able to do so for all but one premium series when testing for single-threshold nonlinearity. Hence, we may conclude, with the exception of the 12 month dollar-yen forward premium, that each premium is decidedly nonlinear and that this linearity is of the single-threshold type. Note that these findings indicate the absence of the inner-regime in model (6.24) yet does not violate our basic premise that the forward premiums are nonlinear in nature.

Table II presents the results of a battery of unit root tests; including the augmented Dickey-Fuller, Enders-Granger (specifying stationary two-regime SETAR and MTAR models under the alternative), Berben-van Dijk (specifying a two-regime CTAR under the alternative), Kapetanios-Shin (specifying a stationary three-regime SETAR model under the alternative) and the lag-augmented double-threshold momentum-TAR models that we proposed as alternative data-generating mechanisms in the unit root tests that we developed in the previous chapter[94,95]. The results generally favor the conclusion that each of the forward

[94] For details on the relative strengths of each of these tests we refer readers to that chapter.
[95] For critical values and additional details with respect to these unit root tests we refer readers to Dickey and Fuller (1979); Enders and Granger (1997); Berben and van Dijk (1999); and Kapetanios and Shin (2002).

Table II
Linear and Nonlinear Unit Root Tests of the Dollar-Sterling, Dollar-Mark and Dollar-Yen Forward Premia

* indicates rejection of the unit root hypothesis at the 5% significance level. Reported Values for the ADF test are pseudo t-statistics while the nonlinear unit root tests specifying single- and double threshold models under the alternative are supremum-Wald statistics. In all cases the tests were performed on "demeaned" series. The ADF, Enders-Granger and van Tol-Wolff tests allow a lag-augmented data generating mechanism with L set to 3. The van Tol-Wolff tests additionally have a maximum threshold delay value of 3. The trimming parameters and the minimal fraction of observations in each regime are set to 15%.

	Linear	Single-Threshold Models			Double-Threshold Models		
Horizon	ADF (AR)	Enders-Granger (SETAR)	Enders-Granger (MTAR)	Berben – van Dijk (CTAR)	Kapetanios-Shin (SETAR)	Van Tol-Wolff (EQ-MTAR)	Van Tol-Wolff (Band-MTAR)
Unit Root Tests of the Forward Premia: Dollar-Sterling							
1-month	-5.390*	21.257*	28.677*	56.441*	114.194*	96.886*	98.162*
3-month	-4.060*	12.479*	14.329*	9.628*	15.684*	16.253*	16.549*
6-month	-3.363*	6.242*	12.112*	6.555*	13.648*	16.607*	17.359*
12-month	-3.409*	6.806*	7.953*	7.024*	11.264*	9.237	10.198*
Unit Root Tests of the Forward Premia: Dollar-Mark							
1-month	-4.987*	16.146*	15.464*	28.704*	30.478*	21.955*	32.683*
3-month	-4.098*	31.288*	22.175*	101.600*	97.816*	82.073*	87.282*
6-month	-3.060*	6.873*	11.399*	6.983*	8.226	4.103	8.446
12-month	-3.005*	7.896*	14.033*	7.731*	9.203*	4.374	12.487*
Unit Root Tests of the Forward Premia: Dollar-Yen							
1-month	-3.720*	15.771*	11.395*	11.132*	8.883	6.779	10.391*
3-month	-3.051*	7.056*	6.139*	6.376*	3.892	3.276	3.231
6-month	-2.713	4.895*	9.168*	4.855*	4.124	6.078	9.996*
12-month	-2.405	3.882	6.780*	4.291	4.023	3.541	11.293*

premium series are stationary processes. Additionally, note that the forward premiums tend to become less stationary as the horizon increases.

Given the evidence in Table I that statistically discernable single-threshold nonlinearity is present in each of the series it is appropriate that in judging whether the processes are stationary most weight is applied to the single-threshold tests. There only the 12 month dollar-yen forward premium is unable to cause rejection of the unit root hypothesis; since there is insufficient evidence of nonlinearity the appropriate test is the ADF test, this test indicates an inability to reject the unit root hypothesis[96].

Given the strong evidence of single-threshold nonlinearity in the forward premiums Balke and Fomby (1997) suggest a two-step approach for examining threshold cointegration utilizing standard time-series methods, which they argue work "reasonably well". Pippenger and Goering (1993, 2000) have, however, poignantly illustrated the detrimental effect that the presence of thresholds, in the equilibrium reversion process, has on both linear unit root and cointegration tests. For the bivariate case Seo and Hansen (2002) and Seo (2003) propose algorithms for the determination of maximum likelihood estimation of a complete threshold cointegration model, including the cointegrating vector. Unfortunately, no test exists for the

[96] The implications of this finding in the out-of-sample forecasting exercises is very strong, generating forecast errors that, in magnitude, far exceed those of a simple (driftless) random walk.

multivariate case and as such we assume a known matrix of cointegrating vectors. Implicitly, the linearity and unit root tests have assumed the existence of the following cointegration matrix

$$\boldsymbol{\beta} = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \end{bmatrix} \tag{6.30}$$

Clarida and Taylor (1997) are unable to reject the set of economically motivated restrictions implied by (6.30); based on a likelihood ratio test of the hypothesis that exactly four linearly independent forward premiums comprise the basis of the cointegration space[97]. Hence, we are in position whereby we assume support for the empirical framework set forth by the authors, by means of assumption (6.30), and have shown overwhelming evidence that the cointegrating residuals, or forward premiums, are well defined as single-threshold SETAR models and are stationary; with the exception of a single forward premium series. Resultantly, we are in a position to estimate the multivariate TVECM these findings imply.

6.4.4 Multivariate Threshold Vector Error Correction Model Results

Since we have found overwhelming support for the hypothesis that equilibrium reversion in each series of forward premiums is asymmetric we desire to incorporate these findings into a model of the form represented by (6.5). Tables IV, V and VI report model estimates[98] when a TAR model of the following form is estimated on the forward premiums in the first-step (Threshold estimates and Box-Ljung statistics are reported at the top of each table)

[97] Our reassessment of their findings in Chapter 3 does however indicate that these findings are not as strong as they claim.

[98] Note that we report the full set of parameter estimates. Firstly, imposing the restriction that the coefficient matrices on all matrices other than the first-order matrix did not result in a significant change in the maximized likelihood; sequential application of LR tests having led to this conclusion. Secondly, we found that removing all insignificant coefficient estimates led to models that appeared rather arbitrary while repetitive removal of the "least" significant regressor tended to destroy the model in it's entirety; this left the model for the UK without any right-side variables, for example – implying that the model degenerated into a random walk model – despite the appeal of this development we find it uninteresting to allow this to occur and hence continue to leave all regressors in the model. Note that Clarida and Taylor (1997) found most regressors to be significant at the 5% level.

Table IV
FIML Error Correction Model for the Five-Variable System: Dollar-Sterling

Sample period is 1977:1 to 1990:26. The Q-statistics are Box-Ljung statistics computed at 13 autocorrelations of the residual series; H is Hosking's multivariate portmanteau statistic computed at 13 autocorrelations; All distributions are distributed as central χ^2 under the null hypothesis, with the degrees of freedom indicated. Figures in parentheses are marginal significance levels.

SETAR Model Parameter Estimates of the Forward Premiums

	$s_t - f_{4,t}$	$s_t - f_{13,t}$	$s_t - f_{26,t}$	$s_t - f_{52,t}$
$\hat{\theta}$	-0.0016	-0.0035	0.0122	0.0257
$Q(13)$	154.830 (0.000)	27.902 (0.009)	44.950 (0.000)	24.028 (0.031)

Multivariate Threshold Vector Error Correction Model Estimates

	Explanatory Variables	Model for Δs_t Coeff.	t	Model for $\Delta f_{4,t}$ Coeff.	t	Model for $\Delta f_{13,t}$ Coeff.	t	Model for $\Delta f_{26,t}$ Coeff.	t	Model for $\Delta f_{52,t}$ Coeff.	t
	Δs_{t-1}	-0.312	-0.533	0.415	0.709	-0.287	-0.487	-0.540	-0.902	-0.702	-1.133
	$\Delta f_{4,t-1}$	0.548	0.850	-0.523	-0.811	0.733	1.128	0.784	1.189	0.728	1.066
	$\Delta f_{13,t-1}$	0.462	0.619	0.644	0.863	-0.164	-0.218	0.517	0.677	0.416	0.527
	$\Delta f_{26,t-1}$	-1.295	-1.828	-1.120	-1.582	-0.787	-1.104	-1.524	-2.105	-0.816	-1.089
	$\Delta f_{52,t-1}$	0.579	1.737	0.562	1.686	0.483	1.438	0.741	2.175	0.355	1.006
Lower Regime	$(s-f_4)_{t-1}$	-0.970	-0.622	-0.380	-0.244	-1.256	-0.799	-1.485	-0.930	-1.480	-0.896
Lower Regime	$(s-f_{13})_{t-1}$	0.099	0.105	-0.103	-0.109	0.601	0.633	0.112	0.116	0.382	0.383
Lower Regime	$(s-f_{26})_{t-1}$	0.172	0.203	0.145	0.171	-0.044	-0.052	0.556	0.640	0.243	0.271
Lower Regime	$(s-f_{52})_{t-1}$	0.015	0.053	0.024	0.085	0.023	0.082	-0.130	-0.456	-0.016	-0.056
Upper Regime	$(s-f_4)_{t-1}$	0.779	0.531	1.682	1.147	0.878	0.594	0.905	0.604	0.894	0.576
Upper Regime	$(s-f_{13})_{t-1}$	-0.187	-0.182	-0.494	-0.480	0.120	0.116	-0.525	-0.499	-0.328	-0.302
Upper Regime	$(s-f_{26})_{t-1}$	0.553	0.588	0.661	0.704	0.680	0.718	1.428	1.487	1.301	1.309
Upper Regime	$(s-f_{52})_{t-1}$	-0.339	-0.900	-0.390	-1.035	-0.491	-1.293	-0.714	-1.853	-0.683	-1.714
	Constant	0.000	0.056	0.000	-0.007	-0.003	-0.417	0.002	0.275	0.001	0.067
		$Q(13) = 9.63$ (0.72)		$Q(13) = 9.62$ (0.73) $H(325) = 533.07$ (0.00)		$Q(13) = 10.35$ (0.67)		$Q(13) = 8.90$ (0.78) $\chi^2 = 16.96$ (1.00)		$Q(13) = 9.146$ (0.762)	

$$\left(s_t - f_{h(j)}\right)_t = \begin{cases} \phi_1^{(2)}\left(\left(s_t - f_{h(j)}\right)_{t-1} - \theta\right) + \varepsilon_t & \text{if} \quad f_{h(j),t-1} > \theta \\ \phi_1^{(1)}\left(\left(s_t - f_{h(j)}\right)_{t-1} - \theta\right) + \varepsilon_t & \text{if} \quad f_{h(j),t-1} \le \theta \end{cases} \tag{6.31}$$

here the threshold functions as the attractor. Also, we have restricted the lag-order and threshold delay variables to 1, for reasons of simplicity. Subsequently, conditional on the estimates of the thresholds for each forward premium series we construct the indicator matrix of (6.16) and estimate a model of the form

$$\Delta \mathbf{y}_t = \boldsymbol{\mu} + \boldsymbol{\alpha}^{(1)}\mathbf{I}^{(1)} \circ \left(s - f_{h(j)}\right)_{t-1}^{(1)} + \boldsymbol{\alpha}^{(2)}\mathbf{I}^{(2)} \circ \left(s - f_{h(j)}\right)_{t-1}^{(2)} + \boldsymbol{\Psi}_1 \Delta \mathbf{y}_{t-1} + \boldsymbol{\eta}_t \tag{6.32}$$

Table V
FIML Error Correction Model for the Five-Variable System: Dollar-Mark

Sample period is 1977:1 to 1990:26. The Q-statistics are Box-Ljung statistics computed at 13 autocorrelations of the residual series; H is Hosking's multivariate portmanteau statistic computed at 13 autocorrelations; All distributions are distributed as central χ^2 under the null hypothesis, with the degrees of freedom indicated. Figures in parentheses are marginal significance levels.

SETAR Model Parameter Estimates of the Forward Premiums

	$s_t - f_{4,t}$	$s_t - f_{13,t}$	$s_t - f_{26,t}$	$s_t - f_{52,t}$
$\hat{\theta}$	-0.0048	-0.0146	-0.0272	-0.0475
$Q(13)$	143.590 (0.000)	97.150 (0.000)	30.930 (0.003)	46.678 (0.000)

Multivariate Threshold Vector Error Correction Model Estimates

		Model for Δs_t		Model for $\Delta f_{4,t}$		Model for $\Delta f_{13,t}$		Model for $\Delta f_{26,t}$		Model for $\Delta f_{52,t}$	
	Explanatory Variables	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
	Δs_{t-1}	-0.796	-0.829	-0.416	-0.434	-1.241	-1.303	-0.546	-0.572	-0.708	-0.735
	$\Delta f_{4,t-1}$	0.993	1.013	0.542	0.553	2.021	2.078	0.757	0.777	0.719	0.730
	$\Delta f_{13,t-1}$	-0.148	-0.364	-0.103	-0.254	-1.055	-2.616	-0.040	-0.098	-0.023	-0.057
	$\Delta f_{26,t-1}$	0.479	0.832	0.470	0.818	0.802	1.405	-0.004	-0.008	0.482	0.834
	$\Delta f_{52,t-1}$	-0.510	-1.333	-0.472	-1.237	-0.516	-1.360	-0.159	-0.418	-0.457	-1.190
Lower Regime	$(s-f_4)_{t-1}$	1.427	0.658	1.993	0.921	2.341	1.088	1.344	0.624	1.458	0.670
Lower Regime	$(s-f_{13})_{t-1}$	2.676	1.873	2.522	1.767	3.334	2.352	2.044	1.440	1.985	1.385
Lower Regime	$(s-f_{26})_{t-1}$	-1.345	-1.433	-1.214	-1.296	-1.83	-1.961	-0.724	-0.777	-0.670	-0.711
Lower Regime	$(s-f_{52})_{t-1}$	0.188	0.566	0.144	0.434	0.266	0.806	0.036	0.109	0.068	0.205
Upper Regime	$(s-f_4)_{t-1}$	1.267	0.749	1.557	0.922	0.563	0.335	1.105	0.657	1.159	0.683
Upper Regime	$(s-f_{13})_{t-1}$	-1.436	-1.447	-1.476	-1.489	-0.486	-0.494	-1.293	-1.312	-1.142	-1.147
Upper Regime	$(s-f_{26})_{t-1}$	0.600	0.865	0.537	0.776	0.024	0.035	0.627	0.910	0.339	0.487
Upper Regime	$(s-f_{52})_{t-1}$	0.040	0.162	0.601	0.243	0.160	0.651	0.010	0.039	0.125	0.501
	Constant	0.000	0.171	0.000	0.204	0.000	0.167	0.000	-0.412	0.000	0.183
		$Q(13) = 13.75$ (0.39)		$Q(13) = 13.95$ (0.38) $H(325) = 457.92$ (0.00)		$Q(13) = 13.21$ (0.43)		$Q(13) = 13.50$ (0.41) $\chi^2 = 88.57$ (0.00)		$Q(13) = 13.39$ (0.418)	

There are two reasons for including just a single lagged difference: the first relates to maintaining enough degrees of freedom for accurate model parameter estimation while the second relates to the fact that we intend to juxtapose our findings with that of Clarida and Taylor (1997) who, due to the instability of the maximum-likelihood algorithm that they employ, estimate a "linear" VECM with only a single lagged-difference; the authors cite their forecasting results as evidence that a single lag is adequate.

Table IV presents results for the dollar-sterling system. The first interesting finding is that the likelihood ratio test of the linearity restriction is unable to reject the null. This is curious as the linearity tests of Table I strongly rejected that the forward premiums are linear AR

Table VI
FIML Error Correction Model for the Five-Variable System: Dollar-Yen

Sample period is 1977:1 to 1990:26. The Q-statistics are Box-Ljung statistics computed at 13 autocorrelations of the residual series; H is Hosking's multivariate portmanteau statistic computed at 13 autocorrelations; All distributions are distributed as central χ^2 under the null hypothesis, with the degrees of freedom indicated. Figures in parentheses are marginal significance levels.

SETAR Model Parameter Estimates of the Forward Premiums

	$s_t - f_{4,t}$	$s_t - f_{13,t}$	$s_t - f_{26,t}$	$s_t - f_{52,t}$
$\hat{\theta}$	-0.0014	-0.0167	-0.0322	-0.0490
$Q(13)$	35.678 (0.001)	13.289 (0.426)	23.085 (0.041)	18.070 (0.155)

Multivariate Threshold Vector Error Correction Model Estimates

		Model for Δs_t		Model for $\Delta f_{4,t}$		Model for $\Delta f_{13,t}$		Model for $\Delta f_{26,t}$		Model for $\Delta f_{52,t}$	
	Explanatory Variables	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t	Coeff.	t
	Δs_{t-1}	-1.021	-1.158	-0.632	-0.716	-0.921	-1.046	-0.974	-1.103	-0.600	-0.669
	$\Delta f_{4,t-1}$	2.507	2.088	1.938	1.613	2.589	2.161	2.334	1.940	1.942	1.590
	$\Delta f_{13,t-1}$	-1.340	-1.403	-1.237	-1.294	-1.849	-1.940	-1.217	-1.271	-1.746	-1.797
	$\Delta f_{26,t-1}$	0.461	0.647	0.552	0.774	0.781	1.098	0.361	0.505	1.172	1.615
	$\Delta f_{52,t-1}$	-0.545	-1.813	-0.556	-1.846	-0.540	-1.780	-0.447	-1.485	-0.716	-2.339
Lower Regime	$(s-f_4)_{t-1}$	-0.094	-0.074	0.366	0.288	-0.300	-0.237	-0.234	-0.184	0.112	0.087
Lower Regime	$(s-f_{13})_{t-1}$	0.998	0.682	0.695	0.475	1.237	0.847	0.556	0.379	0.721	0.485
Lower Regime	$(s-f_{26})_{t-1}$	-1.175	-1.206	-1.112	-1.142	-1.287	-1.325	-0.810	-0.831	-1.224	-1.236
Lower Regime	$(s-f_{52})_{t-1}$	0.297	1.378	0.319	1.478	0.336	1.562	0.277	1.280	0.434	1.979
Upper Regime	$(s-f_4)_{t-1}$	-1.287	-0.718	-0.613	-0.342	-1.155	-0.645	-0.898	-0.499	-0.970	-0.532
Upper Regime	$(s-f_{13})_{t-1}$	1.266	1.110	1.187	1.041	1.764	1.550	1.307	1.144	1.371	1.182
Upper Regime	$(s-f_{26})_{t-1}$	-0.491	-0.617	-0.571	-0.717	-0.788	-0.992	-0.437	-0.548	-0.775	-0.958
Upper Regime	$(s-f_{52})_{t-1}$	0.031	0.142	0.048	0.220	0.068	0.316	-0.004	-0.017	0.143	0.649
	Constant	-0.002	-0.277	0.001	0.083	-0.002	-0.350	-0.003	-0.407	0.001	0.059
		$Q(13)=12.48$ (0.49)		$Q(13)=12.77$ (0.47) $H(325)=403.00$ (0.00)		$Q(13)=12.90$ (0.46)		$Q(13)=12.77$ (0.466) $\chi^2=45.12$ (0.267)		$Q(13)=14.45$ (0.34)	

processes, further evidence being provided by the Q-statistics of the SETAR estimates at the top of the table. One possible explanation could be that the nonlinearity in the forward premiums is mitigated by the inclusion of lagged differences. There is, otherwise, no clearly discernable pattern in the estimated coefficients. The linearity hypothesis is firmly rejected for the dollar-mark and dollar-yen systems where, additionally, the different model adequacy measures indicate that the estimated models are well specified.

6.4.5 Out-of-Sample Forecasting Results

Utilizing models (6.31) and (6.32) we generate dynamic forecasts of the spot exchange rate at 1-, 3-, 6-, and 12-month horizons over the period 1990:27 through 1993:52. In assessing the

out-of-sample forecasting performance of these models we compute similar forecasts using a "linear" VECM, including a single lagged-difference as a regressor; an unrestricted forth order VAR model; a driftless random walk model; a forward premium regression; and a standard forward rate model. While the RMSE and MAE of the multivariate threshold VECM is presented in level that of the alternative forecasting models is presented as a ratio. The ratio is computed by dividing the RMSE/MAE of model (6.32) by that of the alternative model. Hence, values below unity indicate the relative superiority of the approach that we advocate.

The results, which are presented in Tables VI, VII and VIII, show that in two of the three systems model (6.32) out-predicts the "linear" VECM by up to approximately 9%. The proposed model is unable to systematically out-predict the (driftless) random walk model; despite producing lower forecast errors at the 1-month horizon for the dollar-yen system the model tends to veer off strongly at longer forecast horizons, having a ratio of 36 and 38 at the 52-week horizon. Hence, we find strong evidence that not only is the linear cointegration framework adopted by Clarida and Taylor (1997) severely misspecified, due to the clear presence of a single-threshold boundary in the forward premiums, but also that their claims of out-predicting the random walk model at various horizons is overstated and possibly a spurious result despite the elegance of their empirical framework.

Table VI
Results of the Forecasting Exercises: Dollar-Sterling

Notes: Forecast period is 1990:27 to 1993:52. For the VECM the RMSE and the MAE is expressed in levels. For the alternative forecasts, the RMSE or the MAE is expressed as the inverse of its ratio to the corresponding figure for the VECM. Thus a figure less than 1 indicates the relative superior performance by the VECM.

	TVECM (level)	VECM (ratio)	VAR (ratio)	Random Walk (ratio)	Forward Premium Regression (ratio)	Forward Rate (ratio)
RMSE						
4-week horizon	0.0402	1.007	1.007	1.036	0.997	1.039
13-week horizon	0.0860	0.994	0.966	1.143	0.997	1.170
26-week horizon	0.1280	1.012	0.965	1.245	0.940	1.323
52-week horizon	0.4903	2.593	0.431	4.222	2.367	4.947
MAE						
4-week horizon	0.0298	1.015	1.016	1.032	1.011	1.033
13-week horizon	0.0637	1.005	0.970	1.161	1.016	1.181
26-week horizon	0.0911	1.020	0.913	1.295	0.946	1.371
52-week horizon	0.1875	1.156	0.247	2.189	1.009	2.536

Table VII
Results of the Forecasting Exercises: Dollar-Mark

Notes: Forecast period is 1990:27 to 1993:52. For the VECM the RMSE and the MAE is expressed in levels. For the alternative forecasts, the RMSE or the MAE is expressed as the inverse of its ratio to the corresponding figure for the VECM. Thus a figure less than 1 indicates the relative superior performance by the VECM.

	TVECM (level)	VECM (ratio)	VAR (ratio)	Random Walk (ratio)	Forward Premium Regression (ratio)	Forward Rate (ratio)
RMSE						
4-week horizon	0.0350	0.998	1.008	1.026	0.975	1.027
13-week horizon	0.0780	0.990	1.015	1.183	0.971	1.194
26-week horizon	0.1115	0.974	1.043	1.331	0.917	1.357
52-week horizon	0.3051	0.950	1.877	3.732	1.867	3.921
MAE						
4-week horizon	0.0274	0.998	1.007	1.016	0.981	1.017
13-week horizon	0.0625	0.995	1.022	1.178	0.991	1.191
26-week horizon	0.0856	0.979	1.042	1.372	0.902	1.426
52-week horizon	0.2303	0.911	1.670	3.987	1.586	4.111

Table VIII
Results of the Forecasting Exercises: Dollar-Yen

Notes: Forecast period is 1990:27 to 1993:52. For the VECM the RMSE and the MAE is expressed in levels. For the alternative forecasts, the RMSE or the MAE is expressed as the inverse of its ratio to the corresponding figure for the VECM. Thus a figure less than 1 indicates the relative superior performance by the VECM.

	TVECM (level)	VECM (ratio)	VAR (ratio)	Random Walk (ratio)	Forward Premium Regression (ratio)	Forward Rate (ratio)
RMSE						
4-week horizon	0.0268	0.9934	0.9946	0.9738	0.9739	0.9743
13-week horizon	0.0599	0.9758	1.0704	1.1024	1.0438	1.1036
26-week horizon	0.1255	0.9879	1.7635	1.8125	1.5189	1.7941
52-week horizon	3.156	0.9447	12.711	36.652	26.300	35.189
MAE						
4-week horizon	0.0209	0.9920	0.9928	0.9869	0.9677	0.9882
13-week horizon	0.0483	0.9738	1.0598	1.1415	1.0569	1.1375
26-week horizon	0.0969	0.9738	1.6784	1.8763	1.3920	1.8365
52-week horizon	2.3698	0.9420	12.316	38.710	23.612	36.916

Interestingly, when inspecting graphs of the forward premium series, illustrated in figure 1, we find that the forward premiums of the dollar-mark and dollar-yen systems exhibit behavior during the out-of-sample period that significantly deviates from that of the estimation period. These observations make the findings of Clarida and Taylor (1997) all the more surprising.

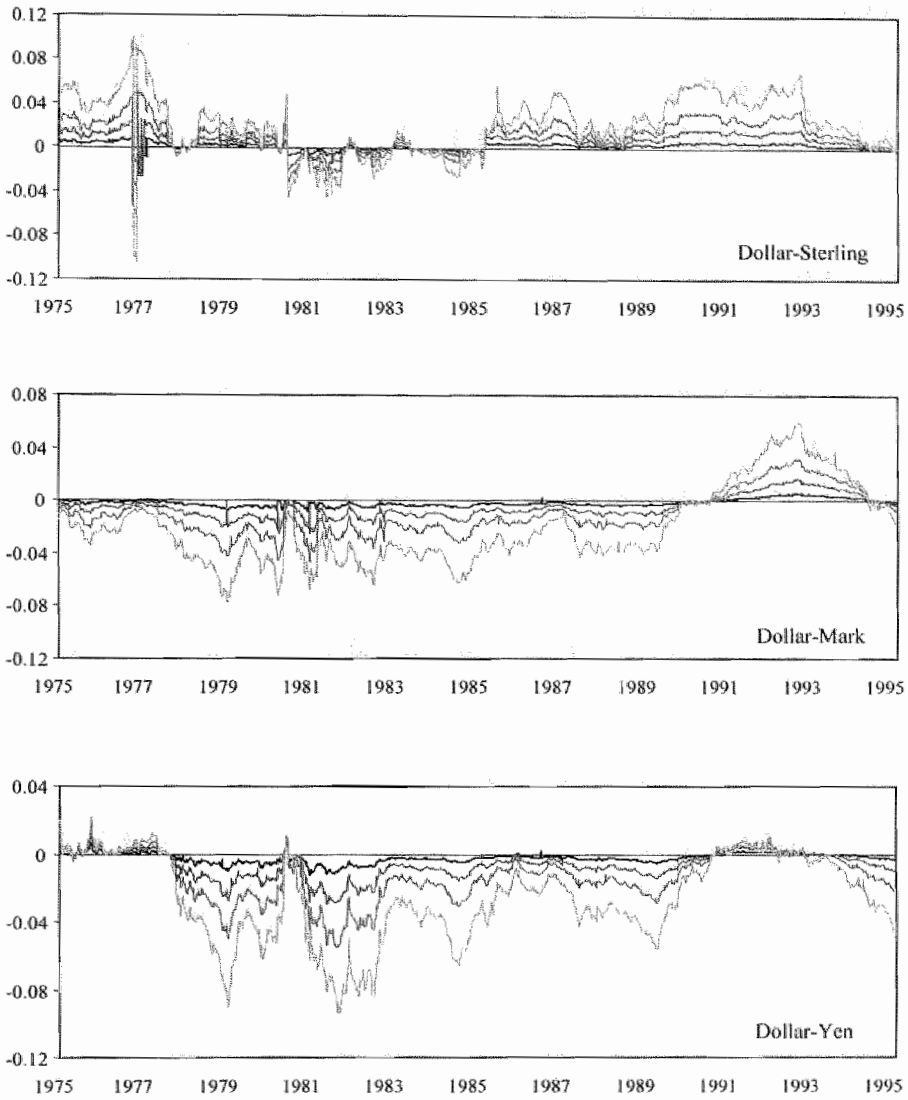

Figure 1. Dollar-Sterling, Dollar-Mark and Dollar-Yen 1-, 3-, 6- and 12-month Forward premia. The dark solid line is the 1-month forward premium while the 3-, 6- and 12-month forward premiums are represented by increasingly less dark solid lines.

6.5. Conclusion

In this Chapter we formulated a convenient specification of a multivariate threshold vector error correction model. The model allows each series of cointegrating residuals to be governed by a nonlinear equilibrium reverting process while obeying long-run equilibrium constraints.

We applied these concepts and constructs to the task of forecasting spot exchange rates utilizing the empirical framework of Clarida and Taylor (1997) in which they show that, under the assumption of stationary deviations from the risk-neutral efficient markets hypothesis, the forward premiums comprise a basis for the cointegration space of a system consisting of spot and forward foreign exchange rates. We were able to establish that the forward premiums exhibit a strong tendency towards equilibrium reversion and that this reversion is asymmetric in nature. Combining this result with the findings of Pippenger and

Goering (2000), who show that standard cointegration tests lack power when the disequilibrium process displays asymmetric reversion, forms a strong argument against the overly-supportive claims of Clarida and Taylor (1997), essentially annulling their findings. Out-of-Sample forecasting exercises illustrated the relative superiority of the multivariate threshold VECM over the "linear" VECM yet, despite this improvement in forecasting performance, we remain incapable of systematically out-predicting a simple (driftless) random walk model. Suggestions for future research include developing procedures capable of incorporating asymmetric equilibrium reversion of the cointegrating residuals when estimating the matrix of cointegrating vectors.

7

Summary and Concluding Remarks

In his Ph.D. thesis *Théorie de la Spéculation* the French mathematician Louis Bachelier, recognized internationally as the father of financial mathematics, proposed that fluctuations in asset prices could be viewed as a random walk. Bachelier assumed that many of these fluctuations followed a Gaussian probability distribution; implicitly presupposing the absence of extreme events. However, it is exactly these extreme events that are so interesting; especially from a risk management perspective. Curiously, however, as has been evinced in this dissertation, the random walk model is still the preeminent benchmark of financial forecasting performance, especially so since the noteworthy results of Meese and Rogoff (1983) in the field of foreign exchange forecasting. Given this setting it is particularly surprising when a, fairly straightforward, study of the information content of foreign exchange forward premiums delivers forecasting results that significantly outperform the random walk benchmark. Chapter 3 details a re-examination of the empirical findings of the article "The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors", by Clarida and Taylor (1997). Unlike the authors we find significantly less support for the empirical framework that they advocate than they claim. The results were not entirely discouraging however hence we sought to incorporate nonlinear

dynamics into their framework. This dissertation details a logical progression from Slutzky's (1927) linear paradigm to the current embrace of nonlinearity in econometric time-series analysis.

Chapter 4 pedantically details an efficient fitting approach in multiple-threshold TAR models. Estimating the parameters of these models is notoriously computationally expensive, and as a result forms a barrier for researchers attempting to investigate the intricacies of the processes that these models represent. The proposed procedure, based on the procedure of Coakley *et al.* (2003), is significantly more computationally tractable than conventional approaches and as such is particularly suited to Monte Carlo simulation analyses. One particularly interesting application of the approach is in the development of unit root testing procedures that allow for threshold nonlinearity under the alternative hypothesis.

Chapter 5, subsequently, details a new unit root testing procedure stipulating a veridical process under the alternative hypothesis. Given the discussion, in the first paragraph above, it is quite apparent that effective models require the capacity to capture both random walk-type behavior in conjunction with deviations thereof. Nonlinear time series models have the capacity to do just that. In this Chapter we focus on the simple TAR model class and are particularly interested in three-regime models allowing for a random walk in the inner-corridor and equilibrium reversion in outer-regimes. Empirical applications of this model type are plentiful: real exchange rate dynamics, purchasing power parity and arbitrage being just a few of the many plausible examples. The models that we stipulate under the alternative hypothesis are stationary double-threshold momentum-TAR models. Summarily, the tests that we develop are superior in terms of power to linear and nonlinear competing tests, in the form of the augmented Dickey-Fuller, Phillips-Perron, Enders-Granger (SETAR and M-TAR) and Berben-van Dijk tests. The size distortion is comparable to that of Berben and van Dijk (1999). By applying the tests to the interest rate differential of UK Eurocurrency interest rates we are able to show that the process is best represented by a Band-MTAR model. An important conclusion that we draw, based on the estimates of a bivariate TVECM and the standard "linear" VECM, is that not incorporating the threshold behavior of the differential will lead to significantly misspecified linear error correction models.

Chapter 6 delineates the logical progression set forth by the nonlinear unit root tests developed in Chapter 5 into the realm of cointegration analysis. In this chapter we develop a multivariate threshold vector error correction model with the ability to allow for independent nonlinear cointegrating residual processes. The framework that we detail is particularly conveniently formulated for estimation purposes. By introducing the notion of an indicator

matrix to differentiate between the various regimes in the set of nonlinear processes we are able to estimate model parameters by OLS. Interestingly, the model defines a combination of linear and nonlinear components; linearity being represented by the set of lagged difference regressors. Hence, the model to some extent entails a *VAR* variant of the constant coefficients method considered by Terui and van Dijk (2002) (see §2.3.4) and allows for TAR-type nonlinearity in its short-run dynamic specification while being subject to long-run equilibrium constraints. We subsequently utilize the model in reassessing the results that we produced in chapter 3. While we are able to improve on the out-of-sample forecasts relative to a linear VECM we are still unable to systematically out-predict a simple (driftless) random walk model. As such we provide strong empirical evidence against the findings of Clarida and Taylor (1997).

Bibliography

Aitken, A.C., 1932, On Interpolation by Iteration of Proportional Parts, without the use of Differences, *Proceedings of the Royal Society Edinburgh* **53**, 54-78.

Anderson, T.W., 1959, On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations, *Annals of Mathematical Statistics*, **30**, 676-687.

Andrews, D.W.K., 1993, Tests for Parameter Instability and Structural Change with Unknown Change Point, *Econometrica,* **61**, 821-856.

Andrews, D.W.K. and W. Ploberger, 1994, Optimal Tests when a Nuisance Parameter is Present only under the Alternative, *Econometrica* **62**, 1383-1414.

Bai, J., and P. Perron, 1998a, Estimating and Testing Linear Models with Multiple Structural Changes, *Econometrica* **66**, 47-78.

Bai, J., and P. Perron, 1998b, Computation and analysis of multiple structural change models, manuscript, Boston University.

Balke, N.S. and T.B. Fomby, 1997, Threshold Cointegration, *International Economic Review* **38**, 627-646.

Bates, J.M. and C.W.J. Granger, 1969, The Combination of Forecasts, *Operational Research Quarterly*, **20**, 451-468.

Baxter, M., 1994, Real Exchange Rates and Real Interest Differentials: Have we missed the Business Cycle Relationship?, *Journal of Monetary Economics*, **33**, 5-37.

Bec, F., Ben Salem, M., and M. Carrasco, 2004, Test for Unit-root Versus Threshold Specification with an Application to the PPP, *Journal of Business and Economics Statistics* **22**, 382–395.

Berben, R. and D. van Dijk, 1999, Unit Root Tests and Asymmetric Adjustment: A Reassessment, *Econometric Institute Research Report EI-9902/A*.

Bessec, M., 2003, The Asymmetric Exchange Rate Dynamics in the EMS: A Time-Varying Threshold Test, *Presented at the International Conference on Policy Modeling* in Istanbul 2003.

Blundell-Wignall, A. and F. Browne, 1991, Increasing Financial market Integration: Real Exchange Rates and Macroeconomic Adjustment, *OECD working paper*.

Bohl, M.T. and P.L. Siklos, 2002, The Bundesbank's inflation Policy and Asymmetric Behavior of the German Term Structure, *Working Paper*.

Boothe, P, and D. Glassman, 1987, Comparing Exchange Rate Forecasting Models: Accuracy versus Proftability, *International Journal of Forecasting* **3**, 65-79.

Boughton, J.M., 1987, Tests of the Performance of Reduced-Form Exchange Rate Models, *Journal of International Economics* **23**, 41-56.

Box, G.E.P. and G.C. Tiao, 1977, A Canonical Analysis of Multiple Time Series, *Biometrika*, **64**, 355-365.

Brock, W., and G. Chamberlain, 1984, Spectral Analysis cannot tell a Macroeconometrician whether his time series came from a stochastic economy or a deterministic economy, SSRI Working Paper 8419, University of Wisconsin Madison.

Brock, W.A., Lakonishok J., and B. LeBaron, 1992, Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *Journal of Finance* **47**, 1731-1764.

Brooks, C., 1996, Testing for Nonlinearity in Daily Pound Exchange Rates, *Applied Financial Economics* **6**, 307-317.

Brooks, C., and M. J. Hinich, 2001, Bicorrelations and Cross-Bicorrelations as Non-linearity Tests and Tools for Exchange Rate Forecasting, *Journal of Forecasting*, **20**, 181-196.

Campbell, J.Y., A.W. Lo, and A.C. MacKinley, 1997, The Econometrics of Financial Markets, Princeton University Press, Princeton.

Caner, M. and B.E. Hansen, 2001, Threshold Autoregression with a Unit Root, *Econometrica* **69** (6), 1555-1596.

Chan, K.S., 1993, Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *The Annals of Statistics* **21**, no.1, 520-533.

Chan, K.S. and H. Tong, 1985, On the Use of the Deterministic Lyapunov Function for the Ergodicity of Stochastic Difference Equations, *Advances in Applied Probability*, **17**, 666-678.

Chan, K.S. and H. Tong, 1986, On Estimating Thresholds in Autoregressive Models, *Journal of Time Series Analysis* **7**, 178-190.

Chan, K.S. and R.S. Tsay, 1998, Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model, *Biometrika*, **85**, 413-426.

Chan, K.S., J.D. Petrucelli, H. Tong and S.W. Woolford, 1985, A multiple threshold AR(1) model, *Journal of Applied Probability*, **22**, 267-279.

Chernick, M.R., 1999, Bootstrap methods: A practitioner's guide, Wiley Series in Probability and Statistics, A Wiley Interscience Publication.

Christiano, L.J., 1992, Searching for a Break in GNP, *Journal of Business and Economic Statistics* **10**, 237- 250.

Clarida, R.H., and M.P. Taylor, 1997, The Term Structure of Forward Exchange Premiums and the Forecastability of Spot Exchange Rates: Correcting the Errors, *The Review of Economics and Statistics* **3**, 353-362.

Clements, P.C. and D.F. Hendry, 2000, Forecasting Economic Time Series, Cambridge UK, Cambridge University Press.

Clyde, W.C., and C.L. Osler, 1997, Charting: Chaos Theory in Disguise?, *Journal of Futures Markets* **17** (5), 489-514.

Coakley, J., A.M. Fuertes and M.T. Perez, 2003, Numerical Issues in Threshold Autoregressive Modeling of Time Series, *Journal of Economic Dynamics and Control* **27**, 2219-2242.

Cook, S., 2003, The Properties of Asymmetric Unit Root Tests in the Presence of Misspecified Asymmetry, *Economics Bulletin* **3**, No. 10, 1-10.

Crane, D.B. and J.R. Crotty, 1967, A Two-stage Forecasting Model: Exponential Smoothing and Multiple Regression, *Management Science*, **13**, B501-507.

Crespo-Cuaresma J., Égert, B., and R. MacDonald, 2004, Nonlinear Exchange Rate Dynamics in Target Zones: A Bumpy Road Toward a Honeymoon, Some Evidence from the ERM,

ERM II, and Selected New EU Member States, *Focus on European Economic Integration* **1** (Oesterreichische NationalBank),46-69.

Davies, R.B., 1977, Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative, *Biometrika* **64**, 247-254.

Davies, R.B., 1987, Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative, *Biometrika* **74**, 33-43.

De Grauwe, P. and D. Decupere, 1992, Psychological Barriers in the Foreign Exchange Market, Centre for Economic Policy Discussion Paper, no. 466, Oct.

Dickey, D. and W.A. Fuller, 1979, Distribution of the Estimates for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* **74**, 427-431.

Dornbusch, R., 1976, The Theory of Flexible Exchange Rate Regimes and Macroeconomic Policy, *Scandinavian journal of Economics*, **78**, 255-275.

Eğecioğlu, O. and A. Srinivasan, 1995, Givens and Householder Reductions for Linear Least Squares on a Cluster of Workstations, *Proceedings of the International Conference on High Performance Computing*, New Delhi, India, 734-739.

Enders, W., 1995, Applied Econometric Time Series," Wiley Series in Probability and Mathematical Statistics, John Wiley and Sons, Inc.

Enders, W., 1999, Improved Critical Values for the Enders-Granger Unit Root Test, *Working Paper 01-02-03*, Department of Economics, Finance and Legal Studies, University of Alabama.

Enders, W. and P.L. Siklos, 1998, Cointegration and Threshold Adjustment, *Department of Economics working paper: Iowa State University.*

Enders, W. and C.W.J Granger, 1998, Unit-Root Tests and Asymmetric Adjustment with an Example using the term structure of Interest Rates, *Journal of Business and Economic Statistics*, **16**, no. 3, 304-311.

Engle R.F. and C.W.J. Granger, 1987, Cointegration and Error Correction: Representations, Estimation and Testing, Econometrica, **55**, 252-276.

Flood, R.P., and N. P. Marion, 1998, Perspectives on the Recent Currency Crisis Literature, *National Bureau of Economic Research Working Paper* No. 6380.

Frankel, J. and K. Froot, 1987a, Short-Term and Long-Term Expectations of the Yen/Dollar Exchange Rate: Evidence from Survey Data, *Journal of the Japanese and International Economies*, **1**, 249-274.

Frankel, J. and K. Froot, 1987b, Using Survey Data to Test Standard Propositions Regarding Exchange Rate Expectations, *American Economic Review*, **77**, no. 1, 133-153.

Frankel, J. and K. Froot, 1990, Chartists, Fundamentalists and the Demand for Dollars, in Anthony Courakis and Mark Taylor, eds., *Private Behavior and Government Policy in Interdependent Economies*, Clarendon Press, Oxford.

Frankel, J.A., and K.A. Froot, 1990, Chartists, Fundamentalists, and Trading in the Foreign Exchange Market, *AEA Papers and Proceedings*, 181-185.

Friedman, M., 1953, The Case for Flexible Exchange Rates, in *Essays in Positive Economics*, University of Chicago, Chicago.

Froot, K., and T. Ito, 1989, On the Consistency of Short-Run and Long-Run Exchange Rate Expectations, *Journal of International Money and Finance*, **8**, no. 4, 487-510.

Georgoutsos, D. and G.P. Kouretas, 2001, Interest Parity, Cointegration and the Term Structure: Testing in an Integrated Framework, *Working Paper.*

Golub, H.G. and J.M. Ortega, 1993, *Scientific Computing: An introduction with Parallel Computing*, Academic Press Inc., San Diego.

Golub H.G. and C.F. van Loan, 1990, Matrix Computations, 2nd edition, The John Hopkins University Press, Baltimore.

Granger, C.W.J., 1981, "Some Properties of Time Series Data and Their Use in Econometric Model Specification, *Journal of Econometrics*, **16**, 121-130.

Granger, C.W.J., 1983, Forecasting White Noise, in *Applied Time Series Analysis of Economic Data*, Proceedings of the Conference on Applied Time Series Analysis of Economic Data, Edited by A. Zellner, U.S. Government Printing Office.

Granger, C.W.J. and P. Newbold, 1974, Spurious Regression in Econometrics, *Journal of Econometrics*, **2**, 111-120.

Granger, C.W.J. and R. Ramanathan, 1984, Improved Methods of Forecasting, *Journal of Forecasting*, **3**, 197-204.

Granger C.W.J., and T. Teräsvirta, 1993, Modeling Nonlinear Economic Relationships, Oxford University Press, Oxford.

Hall, A., H. Anderson, and C.W.J. Granger, 1992, A Cointegration Analysis of Treasury Bill Yields, *The Review of Economics and Statistics* **74**, no. 1, 116-126.

Hamilton, J., 1989, A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* **57**, 357-384.

Hamilton, J.D., 1994, Time Series Analysis, New Jersey, Princeton University Press.

Hansen, B.E., 1996, Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, *Econometrica* **64**, 413-430.

Hansen, B.E., 1997, Inference in TAR Models, *Studies in Nonlinear Dynamics and Econometrics* **2** (1),1-14.

Hansen, B.E., 1999, Testing for Linearity, *Journal of Economic Surveys* **13**, No. 5, 551-576.

Hansen, B.E., 2005, Challenges for Econometric Model Selection, Forthcoming *Econometric Theory* **21** (1).

Hansen, B.E. and B. Seo, 2002, Testing for Two-Regime Threshold Cointegration in Vector Error Correction Models, Journal of Econometrics, *Journal of Econometrics* **110**, 293-318.

Hargreaves, C., 1994, A Review of Methods of Estimating Cointegrating Relationships, in C. Hargreaves (ed.), Nonstationary Time Series Analysis and Cointegration, Chap.4, Oxford University Press, Oxford.

Harvey, A.C., 1989, Forecasting, structural time series models and the Kalman filter, Cambridge UK. Cambridge University Press.

Harvey, A.C., 1993, Time Series Models, 2nd edn. Harvester Wheatsheaf, Hemel Hempstead.

Hendry, D.F., 1987, Econometric Methodology: A Personal Perspective, in Bewley, T. (ed.), Advances in Econometrics, Vol. 2, Cambridge University Press, Cambridge UK.

Hinich, M.J., and D.M. Patterson, 1985, Evidence of Nonlinearity in Daily Stock Returns, *Journal of Business and Economic Statistics* **3** (1), 69-77.

Hosking, J.R.M., 1980, The Multivariate Portmanteau Statistic, *Journal of the American Statistical Association* **75**, 602-608.

Hsieh, D.A., 1989, Testing for Nonlinear Dependence in Daily Foreign Exchange Rates, *Journal of Business* **62** (3), 339-368.

Hsieh, D.A., 1993, Implications of Nonlinear Dynamics for Financial Risk Management, *Journal of Financial and Quantitative Analysis* **28** (1), 41-64.

Hylleberg, S. and G.E. Mizon, 1989, Cointegration and Error Correction Mechanisms, *Economic Journal* **99**, 113-125.

Inder, B., 1995, Finite Sample Arguments for Appropriate Estimation of Cointegrating Vectors, Monash University.

Ito, T., 1994, Short-Run and Long-Run Expectations of the Yen/Dollar Exchange Rate, *Journal of the Japanese and International Economies*, **8**, no.2, 119-143.

Johansen, S., 1988, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control*, **12**, 231-254.

Johansen, S., 1995, Likelihood-Based Inference in Cointegrated Vector Autoregressive models, New York: Oxford University Press.

Kapetanios, G., 1999, Model Selection in Threshold Models, *Working Paper*, Department of Applied Economics, University of Cambridge.

Kapetanios, G. and Y. Shin, 2002, Unit Root Tests in Three-Regime SETAR Models, *Working Paper*.

Kapetanios, G. and Y. Shin, 2003a, Testing for Nonstationary Long Memory against Nonlinear Ergodic Models, Working Paper no. 500, Department of Economics, Queen Mary University of London.

Kapetanios, G. and Y. Shin, 2003b, GLS Detrending-Based Unit Root Tests in Nonlinear STAR and SETAR Frameworks, ESE Discussion Papers from Edinburgh School of Economics, University of Edinburgh.

Kim, T., S.J. Leybourne, and P. Newbold, 2000, Spurious Rejections by Perron Tests in the Presence of a Break, *Oxford Bulletin of Economics and Statistics* **62**, 3, 433-444.

Krugmen, P.R. , 1991, Target Zones and Exchange Rate Dynamics, *The Quarterly Journal of Economics* **106** (3), 669-682.

Larkin, F.M.,1967, Some Techniques for Rational Interpolation, *The Computer Journal* **10**, Issue 2, 178-187.

Lo, M.C., and E. Zivot, 2001, Threshold Cointegration and Nonlinear Adjustment to the Law of One Price, *Macroeconomic Dynamics* **5**, 533-576.

Lütkepohl, H., 1993, Introduction to Multiple Time Series Analysis, 2nd edition, Springer-Verlag.

Madeiros, M.C., Veiga, A., and M.G.C. Resende, 2002, A Combinatorial Approach to Piecewise Linear Time Series Analysis, *Journal of Computational and Graphical Statistics* **11** (1), 236-258.

Mayfield, E.S., and B. Mizrach, 1992, On Determining the Dimension of Real-time Stock-price Data, *Journal of Business and Economic Statistics* **10** (3), 367-374.

Meese, R.A. and K. Rogoff, 1983a, Empirical Exchange Rates of the Seventies, *Journal of International Economics*, **14**, 3-24.

Meese, R.A. and K. Rogoff, 1983b, The Out-of-Sample Failure of Empirical Exchange Rate models, in J. Frenkel, ed Exchange Rates and International Macroeconomics, University of Chicago Press, Chicago.

Meese, R.A., and K. Rogoff, 1986, Was it real? The Exchange Rate-Interest Differential Relation over the Modern Foating-rate Period, *The Journal of Finance* **43** (4), 933-948.

Meese, R.A. and K. Rogoff, 1988, Was it Real? The Exchange Rate – Interest Differential Relationship over the Modern Floating-Rate Period, *Journal of Finance*, **43-4**, 933-948.

Montañés, A. and A. Reyes, 1998, Effect of a Shift in the Trend Function on Dickey-Fuller Unit Root Tests, *Econometric Theory* **14**, 355-363.

Neely, C., Weller, P., and R. Dittmar, 1997, Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach, *Working Paper 96-006C*, Federal Reserve Bank of St. Louis.

Nelson, C.R. and C.I. Plosser, 1982, Trends and Random Walks in Macroeconomic Time Series, *Journal of Monetary Economics*, **10**, 139-162.

Neville, E.H., 1934, Iterative Interpolation, *Journal of the Indian Mathematical Society* **20**, 87-120.

Obstfeld, M., 1990, Intertemporal Dependence, Impatience and Dynamics, *Journal of Monetary Economics* **26**, 45-75.

Perron, P., 1989, The Great Crash, the Oil Price Shock and the Unit Root Hypothesis, *Econometrica* **57**, 1361-1401.

Perron, P., 1993, The Great Crash, the Oil Price Shock and the Unit Root Hypothesis: Erratum, *Econometrica* **61**, 248-249.

Perron, P., 1994, Trend, Unit Root and Structural Change in Macroeconomic Time Series, in Rao, B.B. (ed.), *Cointegration for the Applied Economist*, Macmillan.

Perron, P. and T.J. Vogelsang, 1992, Nonstationarity and Level Shifts with an Application to Purchasing Power Parity, *Journal of Business and Economic Statistics* **10**, 301-320.

Perron, P. and T.J. Vogelsang, 1993, A note on the Asymptotic Distributions of Unit Root Tests in the Additive Outlier Model with Breaks, *Revista de Econometria* **13**, 181-201.

Petrucelli, J.D., 1986, On the Consistency of the Least Squares Estimator of a Threshold AR(1) Model, *Journal of Time Series Analysis* **7**, 269-278.

Petrucelli, J.D. and N. Davies, 1986, A Portmanteau Test for Self-exciting Threshold Autoregressive-type Nonlinearity in Time Series, *Biometrika* **73**, 687-694.

Phillips, P.C.B. and P.Perron, 1988, Testing for a Unit Root in Time Series Regression," *Biometrika* **75**, 335-346.

Phillips, P.C.B. and B.E. Hansen, 1990, Statistical Inference in Instrumental Variables Regression with I(1) processes, *Review of Economic Studies*, **57**, 99-125.

Phillips, P.C.B. and M. Loretan, 1991, Estimating Long-run Economic Equilibria, *Review of Economic Studies*, **58**, 407-436.

Pippenger, M.K. and G.E. Goering, 1993, A Note on the Empirical Power of Unit Root Tests under Threshold Processes," *Oxford Bulletin Of Economics And Statistics* **55**, 473-481.

Pruitt, S.W. and R.E. White, 1988, The CRISMA Trading System: Who says Technical Analysis can't beat the Market? *Journal of Portfolio Management* **14** (3), 55-58.

Pruitt, S.W. and R.E. White, 1989, Exchange-Traded Options and CRISMA Trading, *Journal of Portfolio Management* **15** (4), 55-56.

Quandt, R., 1960, Tests of the Hypothesis that a linear regression obeys two separate regimes, *Journal of the American Statistical Society* **55**, 324-330.

Rao, M.M., 1961, Consistency and Limit Distributions of Estimators of Parameters in Explosive Stochastic Difference Equations, *Annals of Mathematical Statistics*, **32**, 195-218.

Reinmuth, J.E. and M.P, Geurts, 1976, A Multideterministic Approach to Forecasting, in Makridakis, S. and S.C. Wheelwright eds. Forecasting, Vol. 12, TIMS Studies in the Management Sciences, North-Holland, New York, 203-211.

Saikkonen, P., 1991, Asymptotically Efficient Estimation of Cointegrating Regressions, *Econometric Theory*, **7**, 1-21.

Scheinkman, J.A., and B. LeBaron, 1989, Nonlinear Dynamics and Stock Returns, *Journal of Business* **62** (3), 311-337.

Schlittgen, R., 1997, Fitting of threshold models for time series, Discussion Paper, No. 97-13, University of California, San Diego.

Seo, M., 2003a, Nonlinear Mean Reversion in the Term Structure of Interest Rates, *Journal of Economic Dynamics and Control* **27**, 2243-2265.

Seo, M., 2003b, Bootstrap Testing for the Presence of Threshold Cointegration in a Threshold Vector Error Correction Model, *Unpublished Working Paper.*

Séroul, R., 2000, Programming for Mathematicians, Berlin: Springer-Verlag, pp. 216-262.

Sichel, D.E., 1993, Business Cycle Asymmetry: A Deeper Look, Economic Inquiry, **31**, 224-236.

Slutzky, E., 1927, The Summation of Random Causes as the Source of Cyclic Processes, *Econometrica* **5**, 105-146.

Stock, J., and M. Watson, 1988, Testing for Common Trends, *Journal of the American Statistical Association* **83**.

Stock, J.H. and M.W. Watson, 1994, A Simple Estimator of Cointegrating Vectors in higher order integrated systems, *Econometrica* **61**, 783-820.

Stoer, J., 1961, Algorithmen zur Interpolation mit Rationalen Funktionen, *Numerische Mathematik* **3**, 285-304.

Taylor, A., 2001, Potential Pitfalls for the Purchasing-Power-Parity Puzzle? Sampling and specification biases in Mean-reversion tests of the law of one price, *Econometrica* **69** (2), 473-498.

Taylor, M., and H. Allen, 1992, The use of Technical Analysis in the Foreign Exchange Market, *Journal of International Money and Finance* **11**, 304-314.

Terui, N, and H. K. van Dijk, 2002, Combined forecasts from linear and nonlinear time series models, *International Journal of Forecasting*, **18**, issue 3, 421-438.

Thiele, T.N., 1909, Interpolationsrechnung, B.G. Teubner, Leipzig.

Throop, A., 1993, A Generalized Uncovered Interest Parity Model of Exchange Rates, FRBSF Economic Review, **2**, 3 -16.

Tjøstheim, D., 1990, Non-linear Time Series and Markov Chains, *Advances in Applied Probability* **22**, 587-611.

Tong, H., 1978, On a Threshold Model in Pattern Recognition and Signal Processing, ed. C.H. Chen, Amsterdam: Sijhoff &Noordhoff.

Tong, H., 1983, Threshold Models in Non-linear Time Series Analysis," *Lecture Notes in Statistics* **21**, Springer-Verlag.

Tong, H., 1990, Non-linear time series: A dynamical system approach, Oxford Statistical Science series 6, Clarendon Press Oxford 1990.

Tong, H., 1995, Non-Linear Time Series. A Dynamical System Approach, Clarendon Press, Oxford, first Published 1990.

Tong, H. and K.S. Lim, 1980, Threshold Autoregression, Limit Cycles and Cyclical Data, *Journal of the Royal Statistical Society B*, **42**, 245-292.

Tsay, R.S., 1989, Testing and Modeling Threshold Autoregressive Processes, *Journal of the American Statistical Association* **84**, 231-240.

Tsay, R.A., Testing and Modeling Multivariate Threshold Models," *Unpublished Working Paper*.

Vilasuso, J., and S. Cunningham, 1996, Tests for Nonlinearity in EMS Exchange Rates, *Studies in Nonlinear Dynamics & Econometrics* 1 (3), 155-168.

White, J.S., 1958, The limiting Distribution of the Serial Correlation Coefficient in the Explosive Case, *Annals of Mathematical Statistics*, **29**, 1188-1197.

White, J.S., 1959, The limiting Distribution of the Serial Correlation Coefficient in the Explosive Case II, *Annals of Mathematical Statistics*, **30**, 831-834.

Wilkinson, J.H., 1965, *The Algebraic Eigenvalue problem*, Oxford University Press, Oxford.

Wolff, C.C.P., 1987, Time-Varying Parameters and the Out-of-Sample Forecasting Performance of Structural Exchange Rate Models, *Journal of Business and Economic Statistics* **5**, 87-97.

Wolff, C.C.P., 1988, Exchange Rates, Innovations and Forecasting, *Journal of International Money and Finance* **7**, 49-61.

Yule, G.U., 1926, Why Do We Sometimes Get Nonsense Correlations Between Time Series? A Study in Sampling and the Nature of the Time Series, *Journal of the Royal Statistical Society*, **89**, 1-64.

Samenvatting
(Dutch Summary)

Het eindigen van het naoorlogs Bretton Woods systeem van vaste wisselkoersen, in 1973, bewoog de meeste van de grote geïndustrialiseerde economieën tot het laten zweven van hun wisselkoersen. Onderzoek dat hieruit voortvloeide, heeft zich voornamelijk geconcentreerd op de ontwikkeling en schatting van empirische modellen van zwevende wisselkoersen: hoofdzakelijk met het doel om een gezond fundament te verschaffen voor beslissingen op het gebied van economisch beleid. Zwevende wisselkoersen zijn echter notoir lastig om te voorspellen op korte tot middellange termijn.

Terwijl monetaire modellen voor wisselkoers-determinatie beschikken over enig vermogen voor verklaringen op langere termijn, zijn daarentegen korte horizon voorspellingen, gebaseerd op standaard modellen, gebruik makend van waarneembare macro-economische fundamenten, typisch, gemakkelijk overtroffen door een willekeurige gang model.

Het laatstgenoemde werd duidelijk uiteengezet door Meese en Rogoff (1983a,b), die aantoonden hoe een simpel martingale proces in staat is om meer gecompliceerde structurele modellen te overtreffen – bij het voorspellen van wisselkoersen tot een jaar vooruit - ondanks het verstrekken van *ex post* informatie over toekomstige fundamenten zoals geld en productie.

Gedurende de laatste twee decennia hebben talrijke onderzoekers getracht de relatieve prestatie van structurele wisselkoers modellen vis-à-vis atheoretische tijdreeks modellen te bestuderen. Empirisch bewijs is over het algemeen suggestief van het feit dat structurele

modellen hooguit alleen capabel zijn van marginale verbeteringen, buiten steekproef, voor maandelijkse en driemaandelijkse wisselkoers voorspellingen vergeleken met een willekeurige gang model.

Er is significant bewijs van de wijdverspreide toepassing van technische handelsregels door participanten in de financiële markt. Bewijs dat technische analyse hogere winsten kan genereren als een willekeurige handelsstrategie, indien het ware data genererend process niet lineair is, rechtvaardigt haar toepassing en bevestigt de overtuigingen van vele participanten dat chartist-technieken niet lineare componenten exploiteren.

In deze dissertatie concentreren we ons op niet-lineaire tijdreeks modellen. Terwijl de smooth-transition autoregressive, oftewel STAR, modellen worden ondersteund door een goed ontwikkeld fundament van econometrische theorie, wijst empirisch bewijs echter meer in een richting dat zowel reële als nominale wisselkoersen het gedrag vertonen dat meer overeenkomsten vertoont met de dynamiek zoals beschreven wordt door self-exciting autoregressive, oftewel SETAR, modellen. Als gevolg concentreren we ons op deze specifieke categorie modellen.

Zoals eerder werd aangegeven, is het bijzonder moeilijk om een simpele willekeurige gang model buiten-steekproef te overtreffen in voorspellingsoefeningen voor wisselkoers-processen. Vandaar dat wanneer een studie beweert een willekeurige gang, op uiteenlopende horizonnen te overtreffen, het onvoorwaardelijk met enig scepticisme wordt beschouwd. Hoofdstuk 3 is een aantekening op het werk van Clarida and Taylor (1997) die een raamwerk ontwikkelen voor het voorspellen van toekomstige spot koersen door het winnen van informatie van de termijnstructuur van forward koers premies. Dit doen ze door te tonen dat, terwijl zowel de spot als forward koersen individueel geïntegreerde processen zijn, een economisch verdedigbare combinatie van deze processen stationair is met als gevolg dat lineaire cointegratie-technieken kunnen worden toegepast. Gebruik makend van dezelfde data en methodologie waren we niet in staat om hun bevindingen te bevestigen, d.w.z. we vinden significant minder statistische steun voor het empirisch raamwerk dan de auteurs claimen, zelfs wanneer we rekening houden met potentiële redenen voor de discrepantie. Het empirisch raamwerk waar hun studie om draait is echter interessant en leidt vanzelfsprekend tot het idee om niet-lineaire korte termijn dynamiek in het gecointegreerde systeem toe te voegen; een concept dat door Balke and Fomby (1997) in de literatuur werd geïntroduceerd.

Threshold cointegratie is gebaseerd op de schatting van TAR modellen. Hoofdstuk 4 derhalve presenteert een nieuwe schattingsbenadering, gebaseerd op die van Coakley *et al.* (2003), voor het schatten van parameters van een veelvoudig-threshold TAR model. De studie

presenteert een nieuw algoritme voor het construeren van een coördinatenstelsel dat rekening houdt met zowel typisch opgelegde voorwaarden als met een formulering die noodzaaklijk is voor de simpele implementatie van updating algoritmen. Hierop gebaseerd verdelen we de continu threshold ruimte in een set van niet-overlappende hyperkubussen, waarvan elk opeenvolgend wordt geactiveerd en de RSS functie, dat zoals bewezen door Coakley *et al.* (2003), een rationele functie van een specifieke orde is, wordt onderworpen aan een nieuw ontwikkeld interpolatie-gebaseerde optimalisatie routine. Monte Carlo simulaties wijzen op de relatieve superioriteit van de voorgestelde methode wat schattingstijd betreft. Summier is de schattingsmethode 2, 32, 570 en 14.000 keer zo snel in het schatten van 3-, 4-, 5-, en 6-regiem TAR modellen als de conventionele procedure.

Hoofdstuk 5 beschrijft een directe test procedure van de unit root hypothese; met een stationaire double-threshold - TAR model onder de alternatieve hypothese. Een reden hiervoor is de bevinding van Cook (2003), die aangeeft dat beoefenaars bij het modelleren van een onbekend proces meer waarschijnlijk momentum-TAR dan TAR-type dynamiek observeren. De test is uitgebreid aangezien we zowel een lag-augmented data genererende mechanisme als een reeks waarden voor de threshold – delay variable toestaan. Bovendien staan de kritische waarden die we presenteren, de aanwezigheid van een constante en/of een trend attractor toe. Power en size - tests van de voorgestelde test procedure duiden aan dat de test superieur is aan een relevante reeks van alternatieve tests wat power betreft ondanks dat ze geneigd zijn tot over-verwerping; deze size statistieken zijn echter niet ongewoon. We passen de tests toe in een studie van de termijnstructuur van de V.K. Eurovaluta rentevoet. Daar illustreren we dat de rentevoet differentiaal stationair is en het best omschreven wordt door een double-threshold Band-MTAR model. Gebaseerd op deze bevindingen en de aanwezigheid van een enkel unit root in de rentevoet, schatten we een bivariate threshold vector error correction model, oftewel TVECM, en verschaffen we bewijs dat als het threshold gedrag van de rentevoet differentiaal niet in acht wordt genomen, een lineaire VECM significant wordt misspecificeerd.

Hoofdstuk 6 introduceert een schattingsroutine voor multivariate threshold gecointegreerde systemen. Het is verrassend dat, terwijl recente ontwikkelingen in de literatuur zich hebben geconcentreerd op bivariate threshold gecointegreerde systemen, er weinig aandacht is besteed aan de meer algemene multivariate type. We presenteren een geschikte formulering van een multivariate threshold VECM en suggereren een tweestaps-procedure, dat typisch is voor de simpelere bivariate type, door het schatten van (SE)TAR modellen van de gecointegrerende residu reeks en het conditioneren van parameter

schattingen van de multivariate threshold VECM op deze schattingen. We passen deze constructies toe in een heronderzoek van het empirisch raamwerk van Clarida and Taylor (1997), en de aantekening van het derde hoofdstuk. We waren in staat om bewijs te tonen voor de aanwezigheid van single-threshold non-lineairiteit in de forward premies en de daaropvolgende noodzaak om deze karakteristieken op te nemen in het cointegratie raamwerk, geïmpliceerd door hun empirische afleidingen. Buiten-steekproef voorspellingsoefeningen geven aan dat de multivariate threshold VECM relatief superieur is aan de standaard 'lineaire' VECM. Een belangrijke bevinding is dat ondanks de insluiting van het threshold gedrag of van de asymmetrie in het equilibrium terugkeer-proces we nog steeds niet in staat zijn om het willekeurige-gang model systematisch te overtreffen.

Curriculum Vitae

Michel R. van Tol was born on January 7, 1975 in Johannesburg, South Africa. He received his Masters degree in International Business from Maastricht University, the Netherlands, in 2000 having majored in Finance and Statistics. He spent 6 months doing an internship in the North West of Poland for a Dutch software company as well as two years of International Econometrics at Maastricht University. Michel van Tol joined the Department of Finance at Maastricht University as a Ph.D. student in 2000. He joined UBS O'Connor in Chicago as a quantitative strategist in September 2004.